PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [5] :
C12N 15/12, C07K 13/00, A61K 37/02

A2

(11) International Publication Number: WO 94/19463

(43) International Publication Date: 1 September 1994 (01.09.94)

(21) International Application Number: PCT/NZ94/00009

(22) International Filing Date: 16 February 1994 (16.02.94)

(30) Priority Data:
245917 16 February 1993 (16.02.93) NZ

(71) Applicant (*for all designated States except US*): AUCKLAND UNISERVICES LIMITED [NZ/NZ]; c/o The University of Auckland, 58 Symonds Street, Auckland (NZ).

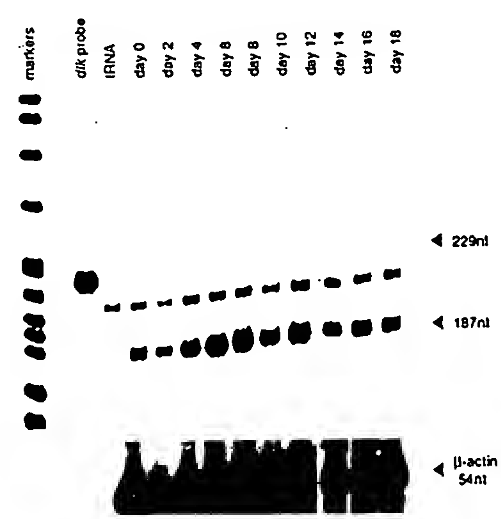
(72) Inventors; and
(75) Inventors/Applicants (*for US only*): CROSIER, Philip, Stanley [NZ/NZ]; 5 Dingle Road, St Heliers, Auckland (NZ). CROSIER, Kathryn, Elizabeth [NZ/NZ]; 5 Dingle Road, St Heliers, Auckland (NZ).

(74) Agents: BENNETT, Michael, R. et al.; A J Park & Son, 6th floor, Huddart Parker Building, Post Office Square, P.O. Box 949, Wellington 6015 (NZ).

(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, ES, FI, GB, GE, HU, JP, KP, KR, KZ, LK, LU, LV, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: DEVELOPMENTAL TYROSINE KINASES AND THEIR LIGANDS



(57) Abstract

The invention relates to mammalian receptor tyrosine kinases designated developmental tyrosine kinases (Dtks). Dtks are expressed in multipotential haematopoietic cells, in embryonic stem cells, in brain tissue and in testis, but are not expressed mature lineage-restricted haematopoietic cells. The invention provides full-sequence Dtks as well as extracellular receptor domains of such Dtks. The invention further provides nucleic acid molecules encoding such Dtks, vectors containing DNA encoding such Dtks, ligands which bind to such Dtks, and methods of therapeutic and/or prophylactic treatment employing either the ligands or extracellular receptor domains.

A8 - 09/090,867

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SI	Slovenia
SK	Slovakia
SN	Senegal
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

DEVELOPMENTAL TYROSINE KINASES AND THEIR LIGANDS

FIELD OF THE INVENTION

The present invention generally relates to protein tyrosine kinase receptors widely expressed by early cells of the haematopoietic system, by cells of the neuronal system in brain tissue, and in testis, ligands for such receptors and nucleic acid molecules encoding such receptors.

BACKGROUND OF THE INVENTION

There are several parallels between the development of the haematopoietic and neuronal systems. In particular, the presence of regulatory protein molecules termed growth factors which recognise and bind to specific cell membrane receptors is a common feature of these two systems. It is possible that shared families of receptors exist that are expressed in both early haematopoietic and neuronal stem cells. In turn, there may be a family of proteins which bind these receptors and function as stem cell growth factors.

The current view of vertebrate haematopoietic ontogeny holds that a succession of pluripotential stem cell migrations originate in the yolk sac blood islands, initially invade the hepatic rudiment, and then the spleen and bone marrow. From the bone marrow, a limited number of multipotential stem cells are laid down during embryogenesis that give rise to a much larger population of developmentally restricted progenitor cells, and ultimately produce the mature cells of at least eight cell lineages. The cells of these lineages are classified as red and white blood cells. The white blood cells contain the mature cells of the lymphoid and myeloid systems. Lymphoid cells contain T and B lymphocytes and are derived from pre-T and pre-B cells, respectively. The myeloid system comprises several cell types known as granulocytes, platelets, monocytes, macrophages, and megakaryocytes. The granulocytes are further divided into neutrophils, eosinophils, basophils and mast cells (see review by Metcalf D. The Molecular Control of Blood Cells, Harvard Univ Press, 1988).

The haematopoietic system functions by precisely controlling the production of cells in the various lineages. Totipotent haematopoietic stem cells have the ability to both self-renew and differentiate. Stem cells undergo a series of differentiation steps leading to increasingly lineage-restricted progenitor cells. The more mature progenitor cells are restricted to production of only one or two lineages. For some time the colony-forming unit-spleen (CFU-S) assay served to operationally define all stem cells. Recent evidence demonstrates heterogeneity within CFU-S, with only a small fraction of CFU-S capable of contributing to long-term repopulation following ablation of the haematopoietic system by irradiation. It is recognised that stem and progenitor populations are not discrete, but represent a continuum of cells from those of high self-renewal capacity and low probability of differentiation to those cells with low self-renewal probability and high commitment to differentiation. When long-term haematopoiesis is investigated at the clonal level, studies have shown that single stem cell clones are sufficient to maintain haematopoiesis over the lifetime of an animal.

The development of the mammalian embryo is governed by interactions between different embryonic cell populations. This process is manifest at the cellular level in the precise temporal and spatial control of proliferation, differentiation and migration. The coordination of these processes may be achieved in part by the action of a family of regulatory molecules termed growth factors. Growth factors can evoke diverse responses in different cell types and may interact with one another synergistically or antagonistically. Their action is complex and most of our current understanding results from *in vitro* experiments. In most instances, haematopoietic growth factor actions defined *in vitro* have been confirmed *in vivo*. In haematopoiesis, some growth factors are lineage-restricted in their action. These include erythropoietin that acts predominantly on red cell development, and granulocyte colony-stimulating factor that's predominant action is on granulocytes. At the other end of the spectrum is interleukin-3 which can act on several target cells such as granulocyte-macrophage progenitors, eosinophils, megakaryocytes, erythroid cells and mast cells. There are no known growth factors that function exclusively on haematopoietic stem cells.

The ligand for c-kit, termed stem cell factor, kit ligand or mast cell growth factor is the product of the Steel (Sl) locus in mice. The factor acts either alone or synergistically with several known growth factors on primitive stem cells. It is believed that this factor is essential for the development of early haematopoietic stem cells, and cells of the erythroid and mast cell lineages.

The stem cell compartment may be viewed as a finely tuned balance between the action of inhibitors and the stimulatory role of cytokines. As with other stem cell systems, haematopoietic stem cells are distributed in a defined spatial manner within adult bones and not in a random, homogeneous mixture of interacting cell types. A concept that underlies the regulation of haematopoietic stem cell development is that these cells reside within a specialised microenvironment, where the regulatory signals act locally. Stromal cells constitute the bone marrow microenvironment.

Embryonic stem cells are permanent cell lines established directly from the inner cell mass of the preimplantation mouse embryo. They retain the ability to participate in normal embryonic development and, following introduction into the blastocyst, generate chimaeric animals that are mosaic in all tissues. Embryonic stem cells are increasingly being used as cellular vectors for experimentally manipulating the mouse genome.

Doetschman has demonstrated that embryonic stem cells can generate primitive erythroid cells in culture (Doetschman et. al. J. Embryol. Exp. Morphol. 87, 27-45; (1985)). This result was achieved by inducing embryonic stem cells to form cystic embryoid bodies in the presence of preselected batches of human cord serum.

In addition to haematopoietic cell development, it has been noted that neurons also arise in differentiating embryonic stem cells. Haematopoietic differentiation in this system occurred infrequently, slowly and was not synchronized. Recently a modified system enabling the differentiation of embryonic stem cells in methylcellulose into multiple haematopoietic lineages has been described by Wiles

and Keller (Development 111, 259-267, (1991)). Using this approach, macrophages, neutrophils, erythroid cells and mast cells develop in a synchronous manner with high frequency in the absence of human cord serum. The development of haematopoiesis from embryonic stem cells in methylcellulose cultures parallels the onset of haematopoiesis in the developing mouse embryo.

An important objective in the field of developmental biology is the identification of genes, the products of which mediate regulatory signals required during embryogenesis. There is compelling evidence that genes encoding receptor tyrosine kinases (RTKs) are involved in early development in vertebrates. The general family of protein tyrosine kinases can be recognised by the presence of several conserved amino acid regions in the catalytic domain. These conserved regions have been summarised by Hanks et al (Science 241, 42-52, (1988)) and by Wilks et al (Proc. Natl. Acad. Sci. USA 86, 1603-1607, (1989)). The receptor for macrophage colony-stimulating factor *c-fms*, which is important in myeloid cell differentiation and placental development is an RTK. The mouse developmental mutation *W* has been shown to involve an RTK. The W locus encodes the c-kit RTK and affects the proliferative and/or migratory properties of primordial germ cells, melanoblasts and haematopoietic stem cells. A recently described RTK termed flk-2, which is related to c-kit, has been isolated using the polymerase chain reaction (PCR) with oligonucleotides to conserved kinase domain motifs. Messenger RNA transcripts for flk-2 are expressed in populations enriched for stem cells and primitive uncommitted progenitor cells, and are absent in mature haematopoietic cells (see Matthews *et al.* Cell 65, 1143-1152, (1991)).

Additional receptor tyrosine kinases expressed on pluripotential haematopoietic stem cells are needed to facilitate the *in vitro* growth of stem cells. The nucleic acid molecules that encode receptor tyrosine kinases expressed by pluripotential stem cells are needed to produce recombinant receptors and ligands.

In vertebrate development, the cells whose descendants give rise to the nervous system are first identified as the neural ectoderm. This forms a tube-like structure beneath the surface of the ectoderm. Following closure of the neural tube some

precursor cells detach from the apical neural tube and form a transient structure called the neural crest. These cells rapidly disperse into the embryo along complex migratory pathways. The proliferating neural crest cells also invade developing tissues such as the skin, gut, and the adrenal gland to form differentiated cell populations within these tissues; eg. melanocytes, enteric neurons and adrenal medullary chromaffin cells.

The diversity of cell types derived from the neural crest poses the problem of how uncommitted embryonic cells acquire particular developmental fates. There are strong parallels between neural crest cell lineage diversification and the process of haematopoiesis. It has been proposed that the earliest neural crest cells should be multipotent and maybe capable of self renewal. Secondly, it should be possible to identify committed progenitors that proliferate symmetrically and are restricted to distinct sublineages and thirdly, there should exist factors which influence the proliferation and/or differentiation of specific types of progenitors (see Anderson Neuron 3, 1-12, (1989)).

Soluble proteins variously termed neurotrophic, growth, and neuronal differentiation factors have been identified that influence the developmental growth, maintenance of function, and plasticity of neuronal populations. These factors have been implicated in the proliferation and differentiation of neurons during embryonic development and in their growth and survival in the adult nervous system. There are a growing number of neurotrophic factors, including nerve growth factor, brain-derived neurotrophic factor, neurotrophin-3 and neurotrophin-4. These molecules constitute a closely related family sharing at least 60% amino acid identity. If the parallel to the haematopoietic system is extended, the range and complexity of cells derived from the neural crest implies that there will be a large number of protein regulators which control this system.

Two different types of receptors have been demonstrated for neurotrophins. One group of these receptors are transmembrane glycoproteins with tyrosine kinase activity encoded by members of the *trk* protooncogene family. It would therefore

be important to isolate additional receptor tyrosine kinases from developing systems such as embryonic stem cells which contain neurons. Ligands for such receptors are required to act inter alia as neurotrophic factors. Nucleic acid molecules encoding the receptors and ligands are needed to produce recombinant receptors and ligands.

It is the object of the present invention to go some way towards fulfilling the above objectives or at least to provide the public with a useful choice.

SUMMARY OF THE INVENTION

The present invention has a number of aspects. In a first aspect, the invention provides a mammalian receptor tyrosine kinase which is a developmental tyrosine kinase (Dtk) and which is expressed in multipotential haematopoietic cells, in embryonic stem cells, in brain tissue and in testis, but which is not expressed in mature lineage-restricted haematopoietic cells.

In a further aspect, the invention provides an extracellular receptor domain of a receptor tyrosine kinase as defined above. In preferred embodiments, this extracellular receptor domain can be bound to a support, or can be in a soluble form.

In still a further aspect, the invention provides a nucleic acid molecule encoding a receptor tyrosine kinase or extracellular receptor domain as defined above. This nucleic acid molecule is preferably DNA.

In yet a further aspect, the invention provides a vector including a DNA molecule as defined above.

In still a further aspect, the invention provides a method of producing a receptor tyrosine kinase comprising the steps of:

(a) culturing a host cell which has been transformed or transfected with a vector as defined above to express the encoded receptor tyrosine kinase or extracellular receptor domain; and

(b) recovering the expressed receptor tyrosine kinase.

As yet an additional aspect, the invention provides a ligand that binds to a receptor tyrosine kinase as defined above.

The ligand can take two forms. In one form, the ligand stimulates the proliferation, differentiation and/or survival of cells which express a receptor tyrosine kinase as defined above (a stimulant ligand).

In the second form, the ligand is antagonistic and at least partially blocks or inhibits the function of a receptor tyrosine kinase as defined above through binding to said receptor (an antagonistic ligand).

In another aspect, the invention provides a method of stimulating the proliferation, differentiation and/or survival of a cell expressing a receptor tyrosine kinase as defined above comprising contacting the cell with a stimulant ligand as defined above.

In yet a further aspect, the invention provides a method of inhibiting the function of a receptor tyrosine kinase as defined above comprising contacting the receptor with an antagonistic ligand as defined above.

In still another aspect, the invention provides a method of treating a disease, syndrome or condition caused or mediated by an excess of a ligand as defined above comprising the step of contacting said excess of said ligand with an effective amount of a receptor tyrosine kinase or an extracellular receptor domain as defined above.

In another aspect, the invention provides a method of extracting a ligand from a medium which may contain said ligand comprising the step of contacting said medium with a receptor tyrosine kinase or with an extracellular receptor domain as defined above.

The invention also provides a method of isolating ligand(s) from a medium which may contain said ligand(s), comprising the steps of:

(a) contacting said medium with an effective amount of a receptor tyrosine kinase or an extracellular domain as defined above;

(b) detecting which ligand(s) bind; and

(c) isolating such bound ligand(s).

While the invention broadly consists in the foregoing, it should be appreciated that it also includes the more specific embodiments detailed in the following description:

DESCRIPTION OF THE FIGURES

Figure 1 shows expression of murine Dtk in embryonic stem (ES) cells and embryoid bodies. RNase protection analysis was performed on total RNA (10 μg) from ESD3 ES cells growing in Leukaemia Inhibitory Factor (LIF) (day 0), or from ES cells maintained in the absence of LIF that were differentiating and developing into cystic embryoid bodies (days 2 to 18). As a control tRNA (10 μg) was also used. The markers were pBR322 digested with *Msp* I. The size of the free murine Dtk probe was 229 nt. A fully protected fragment representing the presence of murine Dtk transcripts was 187 nt in length. The free ß-actin protected fragment is shown in each lane as an RNA loading control.

Figure 2 shows expression of murine Dtk in embryonic mouse tissues. RNase protection analysis was performed on total RNA (10 μg) isolated from E14.5 embryonic tissues of the C57BL/6J mouse strain. Details of the markers, probes and controls are as described for Figure 1.

Figures 3 and 4 show expression of murine Dtk in adult mouse tissues. RNase protection analysis was performed on total RNA (10 μg) isolated from the various tissues of adult C57BL/6J mice. Details of the markers, probes and controls are as described for Figure 1.

Figure 5 shows expression of murine Dtk in murine cell lines. The most abundant expression is in the multipotential cell lines FDC-P1 and DA2, and the mast cell line P815. The majority of other cell lines are lineage-committed, mature haematopoietic cell lines, which have very limited murine Dtk expression. The NIH 3T3 cell line is derived from embryonic fibroblasts and C2C12 is a myoblast cell line.

Figure 6 shows the cDNA and amino acid sequence of murine Dtk.

Figure 7 shows the cDNA and amino acid sequence of human Dtk.

DETAILED DESCRIPTION OF THE INVENTION

A. Receptors

In a first aspect, this invention provides a mammalian receptor protein tyrosine kinase (PTK). The mammal in which the PTK exists may be any mammal, such as a mouse, rat, rabbit or human.

Members of the PTK family are recognised by the conserved amino acid regions in the catalytic domains. Examples of PTK consensus sequences have been provided by Hanks *et al.* (Science 241 42-52 (1988), especially Figure 1 starting at page 46) and by Wilks *et al.* (Proc. Natl. Acad. Sci. USA 86, 1603-1607 (1989), especially Figure 2 on page 1605).

Hanks et al. identify eleven catalytic subdomains containing PTK consensus residues and sequences. The PTKs of the present invention contain most or all of these consensus residues and sequences.

As indicated above, the PTKs of the invention are receptor PTKs and so are also generally referred to as RTKs. Further, as the applicants believe that the RTKs of the invention are involved in mammalian cell development, they are specifically referred to hereinafter as developmental tyrosine kinases (Dtks).

The Dtks of the invention are transmembrane receptor tyrosine kinases whose extracellular domains contain two immunoglobulin-like motifs followed by two fibronectin-type III repeats. RTKs of this structure (Axl(Ufo,Ark)) are already known (Janssen *et al.*, Oncogene 6, 2113-2120 (1991); O'Bryan *et al.*, Mol. Cell. Biol 11, 5016-5031 (1991); Rescigno *et al.* Oncogene 6, 1909-1913 (1991); Faust *et al.* Oncogene 7, 1287-1293 (1992)). The Dtks of the invention are however distinguished from those RTKs having the equivalently structured extracellular domains by their potential function based upon their distribution within the mammalian body.

With regard to this latter feature of the Dtks of the invention, the applicants have conducted experiments to determine the range of cells in which the developmental tyrosine kinases of the invention are expressed. These experiments were specifically performed in relation to murine Dtk but are believed to be illustrative of the expression of all mammalian Dtks of the invention.

A.1 Analysis of Murine Dtk expression

The expression of murine Dtk in a range of embryonic and adult mouse tissues was analyzed by ribonuclease protection analysis, using a probe that encompassed sequences encoding the membrane-proximal portion of the extracellular domain of the receptor.

Materials and Methods

1. Embryonic stem cell culture

The ESD3 embryonic cell line (Doestschman et al., J. Embryol. Exp. Morphol. 87 27-45 (1985)) was maintained on gelatin-coated dishes in Dulbecco's-modified Eagle's medium (DMEM) with additives according to established procedures (Hogan et al., Cold Spring Harbour Laboratory, 1-332 (1986)), in the presence of LIF. Cystic embryoid bodies were established following collagenase treatment of the ES cells and subsequent suspension culture in bacteriological-grade petri dishes in DMEM with additives in the absence of LIF (Wiles and Keller, Development 111, 259-265 (1991)).

2. Fetal liver haematopoietic stem cell enrichment

Low density haematopoietic stem cells were isolated from an E14.5 fetal liver cell suspension using equilibrium density centrifugation on a discontinuous metrizamide gradient according to the method of Visser et al., J. Exp. Med, 59, 1576-1590 (1984). Following this procedure, low density fetal liver cells (p24 < 1.078 g/cm^3) were incubated for 20 minutes on ice in DMEM medium with 5 μg/10^6 cells of AA4 monoclonal antibody (rat IgG_{2b}; McKearn et al., Proc. Natl. Acad. Sci. USA, 82, 7414-7418 (1985)) and washed twice. This antibody has been shown to recognise the most primitive haematopoietic stem cell in fetal liver (Jordan et al., Cell, 61, 953-963 (1990)). The AA4 labelled cells were then incubated on ice for 20 minutes with magnetic beads conjugated with anti-rat IgG antibody as outlined in the manufacturer's protocol (Advanced Magnetics Corp., Cambridge, MA). Following incubation, AA4$^+$ cells were positively-selected on a magnet. Stem cell enrichment was assessed by re-labelling the cells with the AA4 antibody, followed by a second layer antibody staining with goat anti-rat fluorescein isothiocyanate and flow cytometric analysis on a FACS 440 (Becton Dickinson, San Jose, CA).

3. RNA analysis

RNase protection analysis was performed by hybridization of 10 μg of total RNA to RNA probes that encoded sequences of murine Dtk and ß-actin, overnight at 52°C. RNase digestion was performed with RNase T1 (1.75 μg/ml) and RNase A (35 μg/ml) at 37°C for one hour. The reaction was stopped with proteinase K (333 μg/ml) and SDS (0.3%). The products were run on a 6% urea/acrylamide gel and the autoradiograph exposed at -70°C. The probe for analysis of Dtk expression was derived from nucleotides 1158 to 1334 of the Dtk sequence, a segment which encodes the membrane-proximal portion of the extracellular domain, and which had been subcloned into pGEM-4Z. In an RNase protection assay, the free probe yielded a 229 nucleotide (nt) band, and Dtk transcripts protected a fragment of 187 nt. A riboprobe was also constructed from a Sal I-Sma I fragment of human ß-actin. The length of the free ß-actin probe was 132 nt and ß-actin transcripts protected a 54 nt fragment.

Results

1. Embryonic stem cells

Figure 1 demonstrates the expression of Dtk transcripts in both totipotent ES cells growing in LIF (termed day 0), and in differentiating cystic embryoid bodies growing in the absence of LIF for up to 18 days. In this developmental system Dtk is expressed almost uniformly from days 0 to 18, indicated by the presence of a protected 187 nt band for each RNA analyzed. The two bands of approximately 220 nt and 210 nt present in lanes for each RNA sample analyzed are also present in the tRNA lane and are regarded as nonspecific. Of considerable interest with regard to the importance of this receptor in mouse development is the demonstration of Dtk expression in totipotent ES cells. The ES cells from which RNA was extracted for day 0 analysis were selected from cultures, following morphological assessment by phase-contrast microscopy to confirm that they were undifferentiated.

2. Embryonic tissues

Expression of Dtk was detected in total RNA isolated from a wide range of mid-gestational E14.5 embryonic mouse tissues including the brain, eye, thymus, lung, intestine, forelimb, hindlimb and testis (Figure 2). There was limited expression in heart and unfractionated liver.

Figure 2 shows enrichment of Dtk transcripts in E14.5 fetal liver low density $AA4^+$ haematopoietic stem cells. Following density-gradient centrifugation and positive selection, the cells used for RNA analysis were greater than 95% $AA4^+$, as assessed by flow cytometry (data not shown). Dtk expression was also detected in day 14.5 placenta.

3. Adult tissues

In contrast to the widespread expression of Dtk in embryonic tissues, the pattern of expression in adult tissues becomes restricted (Figures 3 and 4). Dtk transcripts were most abundant in brain, esophagus, bladder, testis, and ovary. In brain, expression of Dtk (relative to ß-actin) was more abundant in adult than in embryonic tissue. Adult tissues which contained less abundant, but detectable

transcripts were lung, and regions of the gastrointestinal tract including the stomach and both the small and large intestine. Tissues in which Dtk transcripts were undetectable oı expressed at extremely low levels included the salivary gland, thymus, heart, liver, skeletal muscle, kidney, spleen, bone marrow, adrenal gland and uterus.

4. Murine cell lines

The pattern of expression of Dtk in murine cell lines was analyzed in relation to the following: WEHI-3B, 416B, EL4, SO3, SP2/0, P388D$_1$, P815, FDC-P1, DA2, FDC-P1/IL-2 ras, NIH3T3 and C2C12. The results are shown in Figure 5.

As can be seen from Figure 5, the results are consistent with those above, with the most abundant expression being in the multipotential cell lines FDC-P1 and DA2, and in mast cell line P815. Significant expression is also observed in myoblast cell line C2C12.

In contrast, the remaining cell lines (lineage-restricted mature haematopoietic cell lines) show very limited murine Dtk expression.

From this analysis, the applicants have derived the condition defining the Dtks of the invention - they are expressed in multipotential haematopoietic cells, in totipotent embryonic stem cells, in brain tissue and in testis, but not in mature lineage-restricted haematopoietic cells.

For the purpose of this specification, a multipotential haematopoietic cell is an early haematopoietic cell. Examples of multipotential haematopoietic cells include multipotential factor-dependent cells that have the capacity to proliferate and differentiate into mature haematopoietic cells. In contrast, a mature haematopoietic cell is non self-renewing and has limited ability to give rise to multiple cell lineages. Mature lineage-restricted haematopoietic cells, for the purposes of this specification, are therefore represented by haematopoietic cell lines of the T or B lymphoid lineage or mature myeloid lineages.

The Dtks of the present invention may or may not be expressed in intermediate cells poised between the state of being multipotential and mature.

In terms of brain tissue, the Dtks of the invention are primarily expressed in neuronal cells.

In terms of testis, the Dtks are primarily expressed in the Sertoli cells.

It will of course be appreciated by those persons skilled in this art that the reference to the Dtks of the invention not being expressed in mature lineage-restricted haematopoietic cells is in a biological context and does not mean that there is absolutely no expression of the Dtk in these cells. As is apparent from Figures 1 to 5, what is meant by the phrase "not expressed in mature-lineage restricted haematopoietic cells" is that there is no significant expression of the Dtk in the cell, i.e. that expression is either undetectable or at an extremely low level.

The restricted expression of the Dtks of the invention to cells representative of early multipotential cells, with substantial absence of expression in lineage-restricted cells such as T or B lymphocytes, is consistent with this receptor functioning and transducing signals from the microenvironment to the haematopoietic stem cell compartment. The expression of the Dtk in embryonic stem cells and in some fetal tissues such as brain is also consistent with this receptor and its ligand having a functional role in the specification of cell lineages during embryonic development, including neuronal development. Furthermore, the receptor and its ligand is likely to have a role in the maintenance of function and plasticity in neuronal populations or their derivatives. Finally, the expression of the receptor in adult brain is consistent with the receptor and its ligand having a role in the growth and survival of neurons in the adult nervous system.

The embryonic stem cell and haematopoietic multipotential cell line mRNA for Dtk migrates relative to 28S and 18S ribosomal bands on formaldehyde agarose gels at approximately 4.2 Kb. In adult brain tissues, Dtk mRNA migrates at approximately 4.2 Kb.

The Dtks of the invention can usefully be provided in a number of different forms. These include the Dtk itself, the "mature" form of the Dtk, and the extracellular receptor domain of the Dtk.

The "mature" form of the Dtk of the invention is the Dtk less its native amino-terminus leader or signal sequence, whereas the extracellular receptor domain is the Dtk lacking the transmembrane region and catalytic domain.

The extracellular domain may be identified through commonly recognised criteria of extracellular amino acid sequences. The determination of appropriate criteria is known to those skilled in the art, and has been described, for example by Hopp et al., Proc. Natl. Acad. Sci. USA 78, 3824-3828 (1991); Kyte et al., J. Mol. Biol. 157, 105-132 (1982); Emini, J. Virol 55, 836-839 (1985); Jameson et al. CA BIOS 4, 181-186 (1988); and Karplus et al. Naturwissenschaften 72, 212-213 (1985). Amino acid domains predicted by these criteria to be surface exposed are characteristic of extracellular domains.

The Dtks of the invention or their extracellular receptor domains may be prepared by methods known in the art. Such methods include protein synthesis from individual amino acids as described by Stuart and Young in "Solid Phase Peptide Synthesis", Second Edition, Pierce Chemical Company (1984). It is however preferred that the Dtks and/or their extracellular receptor domains be prepared by recombinant methods as will be detailed hereinafter.

A.2 Specific Dtks of the Invention

A.2.1 Murine Dtk

As is indicated above, a first Dtk of the invention, murine Dtk, has been identified in certain tissues of the mouse. Murine Dtk generally has the nucleic acid and deduced amino acid sequence shown in Figure 6. Figure 6 represents individual amino acid residues as single letters as follows:

Amino Acid	Three-letter abbreviation	One-letter symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Asparagine or aspartic acid	Asx	B
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic Acid	Glu	E
Glutamine or glutamic acid	Glx	Z
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Details of the sequence of murine Dtk are as follows.

<u>Sequence Analysis of the Murine Dtk</u>

Figure 6 shows the 3.919 Kb nucleotide and deduced amino acid sequence for murine Dtk from murine neonatal brain. Within the 5' region, a potential site for translation initiation (-GGAGCATGGGG-) is found within a good Kozak consensus sequence. The first methionine initiates an open reading frame of 874 amino acids. Using the method of von Heijne <u>Sequence Analysis in Molecular Biology</u> 113-117, San Diego, Academic Press (1987), the signal cleavage site is predicted to be between alanine 24 and alanine 25, which specifies a 24 amino

acid hydrophobic leader sequence and a mature receptor tyrosine kinase protein of 850 amino acids. Amino acids AGLK to PHSR form a 386 amino acid extracellular domain. A 25 amino acid hydrophobic region from TSWV to LILL is consistent with that of a transmembrane domain (Fasman and Gilbert, Trends Biochem 15, 89-92 (1990)), while the remaining amino acids ending HSSC comprise the cytoplasmic domain.

The extracellular domain of murine Dtk contains eight consensus sites (NxT or S) for *N*-linked glycosylation, predicting that the mature Dtk protein is glycosylated. Within the extracellular domain, two repeating protein motifs are identifiable. Using the predictive methods of Williams and Barclay, Ann. Rev. Immunol 6, 381-405 (1988), two C-type immunoglobulin (Ig)-like domains are present from amino acids KLMG to GEET (Ig-like domain I) and FFTV to NIKG (Ig-like domain II). The first Ig domain has a structure similar to a C1 domain, while the second Ig domain is more C2-like. Based on the analysis of Petersen *et al.*, Proc. Natl. Acad. Sci. USA 80, 137-141 (1983), there are two fibronectin type III modules present from amino acids PPAA to PYGD (domain I) and from amino acids PFQT to SHDH (domain II).

Analysis of the 439 amino acid cytoplasmic domain sequence of murine Dtk shows many of the motifs which are highly conserved within the catalytic kinase domain of protein tyrosine kinases (Hanks *et al.*, Science 241, 42-52 (1988)). The motifs GKGEFG and VAVK, which function as the Mg^{2+} -ATP binding site (Ullrich and Schlessinger, Cell 61, 203-212 (1990); Cantley et al., Cell 64, 281-302 (1991)), are observed at the start of the kinase domain. Further towards the carboxy-terminus

of Dtk other conserved kinase motifs are identifiable, including the motif IHRDLAARN, the DFG triplet motif and the motifs KWLALES and DVWAFG. Alignment of the kinase domain of Dtk with other protein tyrosine kinase domains including that of Ufo, suggests there is a kinase insert region specified by the amino acids RIGENPFN. There are 12 tyrosine residues within the cytoplasmic domain of Dtk, including two residues located near the C-terminus that are nested within sequences that exhibit strong homology to Src homology 2 (SH2) domain binding sites (Songyang et al., Cell 72, 767-778 (1993)). One of these sequences, EEVYDLM, is a putative binding site for phosphatidylinositol 3-kinase, but lies within the catalytic domain proper and is unlikely to be autophosphorylated. The sequence DPLYINI fulfills criteria for either a Sem5/Grb2 binding site or a phospholipase C-γ binding site (Songyang *et al.*, (1993)) supra, and its position in the C-terminal tail makes it a good candidate for phosphorylation.

In specific aspects, the invention provides murine Dtk, mature murine Dtk and the extracellular receptor domain of murine Dtk.

Murine Dtk has the amino acid sequence given as SEQ ID NO 1.

Mature murine Dtk has the amino acid sequence given as SEQ ID NO 2.

The extracellular receptor domain of murine Dtk has the amino acid sequence given as SEQ ID NO 5.

The invention also includes functional equivalents of murine Dtk, mature murine Dtk and the extracellular receptor domain of murine Dtk as is described hereinafter.

A.2.2 Human Dtk

A second Dtk of the invention has been identified from human tissue. This second receptor is the human homologue of murine Dtk having all of the structural features of murine Dtk.

The nucleic acid and deduced amino acid sequence for this receptor tyrosine kinase, hereinafter called "human Dtk", is shown in Figure 7. Sequence details are as follows.

Sequence Analysis of the Human Dtk

Figure 7 shows the 4.364 Kb nucleotide and deduced amino acid sequence for the human Dtk from human fetal brain. The structural features of human Dtk closely parallel those described for murine Dtk. The signal peptide encompasses amino acids MGRP to ESAA. The mature protein extends from residues AGLK to HSSC. Within the mature protein the extracellular domain is defined by residues AGLK to PHSR, the transmembrane domain by residues TSWV to LILL, and the cytoplasmic domain from residues RKRR to HSSC.

The extracellular domain contains two repeating protein motifs made up of two immunoglobulin domains (KLMG to GGET and FFTV to NLKG), followed by two fibronectin type III modules (LPAA to PYAD and PFQT to SHDR). The

protein tyrosine kinase domain is encompassed by the amino acids LGKG to RMEL within the cytoplasmic domain. The motifs defined within the murine protein tyrosine kinase domain are also identifiable within the human protein tyrosine kinase domain.

Once again, in its specific aspects the invention provides different forms of the Dtk (human Dtk, "mature" human Dtk and the extracellular receptor domain of human Dtk).

Human Dtk has the amino acid sequence given as SEQ ID NO 3.

Mature human Dtk has the amino acid sequence given as SEQ ID NO 4.

The extracellular receptor domain of human Dtk has the amino acid sequence given as SEQ ID NO 6.

Once again, the invention further includes functional equivalents of human Dtk, mature human Dtk and of the extracellular receptor domain of human DTK.

A.2.3 Other Mammalian Dtks

In addition to the murine and human Dtks described above, the invention includes within its scope Dtks of other mammals. Such Dtks are the homologues of both murine and human Dtk and can be readily identified by those persons skilled in the art with reference to the characterising data given above for murine Dtk and human Dtk.

By way of example, one method for identifying other Dtks of the invention involves the formation of a DNA library from a suitable tissue source (such as brain) obtained from the mammal. This library can then be screened to identify DNA coding for homologues to murine Dtk and human Dtk as will be described in more detail below.

B. Nucleic Acid Molecules Encoding the Dtks of the Invention

In another aspect of this invention, the applicants provide nucleic acid molecules encoding the Dtks. These nucleic acid molecules may be DNA (isolated from nature, synthesised or cDNA) or RNA. Most often, the nucleic acid molecules will be DNA.

B.1 Nucleic Acid Molecules Encoding Murine Dtk and Human Dtk

As indicated above, the nucleic acid sequences for murine Dtk and human Dtk have been determined. In specific aspects, the invention therefore provides nucleic acid molecules (in the form of DNA) as follows:

1. A DNA molecule encoding murine Dtk having the nucleotide sequence given as SEQ ID NO 7.

2. A DNA molecule encoding mature murine Dtk having the nucleotide sequence given as SEQ ID NO 8.

3. A DNA molecule encoding the extracellular receptor domain of murine Dtk having the nucleotide sequence given as SEQ ID NO 11.

4. A DNA molecule encoding human Dtk having the nucleotide sequence given as SEQ ID NO 9.

5. A DNA molecule encoding mature human Dtk having the nucleotide sequence given as SEQ ID NO 10.

6. A DNA molecule encoding the extracellular receptor domain of human Dtk having the nucleotide sequence given as SEQ ID NO 12.

The invention also includes within its scope functional equivalents of these DNA molecules.

B.2 Nucleic Acid Molecules Encoding Dtks of other Mammals

It will be appreciated that DNA molecules encoding the functional equivalent homologues of murine Dtk and human Dtk from other mammals are also within the scope of the invention. Such DNA molecules can be readily identified using conventional techniques and with reference to the information contained herein characterising murine Dtk and human Dtk.

By way of generic illustration, DNA molecules encoding homologues of murine Dtk and human Dtk in other mammals can be identified by employing the following general steps:

(a) Formation of a cDNA library:

Total mRNA from a suitable tissue source (such as brain) of the mammal is prepared by standard procedures (Ausubel et al, (Eds),

"Current Protocols in Molecular Biology" Greene Associates/Wiley Interscience, New York (1990)), and cDNA synthesised. A cDNA library is formed (for example in λZAP II).

(b) Library Screening:

The cDNA library formed as above is screened for the presence of cDNA encoding homologues to murine Dtk and human Dtk. Screening will generally employ a DNA hybridisation or amplification step with the probes or primers being selected based upon the already determined sequences of murine and human Dtk.

Most conveniently, the screening procedure will involve DNA amplification using the polymerase chain reaction (PCR) (Saiki et al Science 239, 487 (1988)) with the PCR primers being selected such that highly conserved regions from within the DNA sequence of murine and human Dtk will be within the amplified PCR product.

(c) DNA Isolation and Sequencing:

Clones from the cDNA library which are identified by screening step (b) as containing cDNA encoding homologues to murine and human Dtk are selected, and the size of the cDNA insert sourced from the brain determined. Such clone(s) including a cDNA insert of the appropriate size to code for the full-length Dtk are selected and the cDNA insert isolated. Each isolated cDNA insert is then sequenced using known procedures (for example, using the standard

dideoxy chain-termination method of Sanger et al., Proc. Natl. Acad. Sci. USA 74, 5463-5467 (1977)).

B.3 Genetic Mapping of Murine Dtk and Human Dtk

By way of further characterisation of both murine Dtk and human Dtk, the applicants have performed experiments to establish the chromosomes on which the genes coding for these Dtks are located. Details of these experiments are given below.

Materials and Methods

B.3.1 Fluorescent In Situ Hybridization (FISH)

A partial Sau3A genomic DNA library in λ 2001, prepared from mouse ES cells (Boehm et al., Proc. Natl. Acad. Sci. USA 88, 3927-3931 (1991)), was screened with the 3.525 kb cDNA insert purified from pMo23A using methods previously described (Morris, et al., Blood 76, 1812-1818 (1991)). Of 34 positive clones, two of the most intensely hybridizing, λMo23A-7.1 and λMo23A-8.1, were selected for FISH studies. The pMo23A plasmid, and DNA isolated from bacteriophage clones λMo23A-7.1 and λMo23A-8.1, were biotinylated by nick-translation using biotin-14-dATP (Bethesda Research Laboratories, Gaithersberg, MD). Karyotypically normal, 40,XY, mouse metaphase cells were prepared from ES cells in culture using standard procedures. Fluorescent in situ hybridization and detection procedures were essentially as described (Morris et al., Human Genetics 91, 31-36 (1993)), except that mouse Cot 1 DNA (BRL, final concentration 250ng/μl) was used to suppress repetitive sequences in the two phage DNA probes. Chromosomes were G-banded using DAPI (4',6-diamidino-2-

phenylindoledihydrochoride, Sigma, St Louis, MO) as a counterstain for fluorescence analysis.

B.3.2 Single-strand conformation polymorphism (SSCP)

Primer sequences from the 3' untranslated region of the Dtk cDNA used for genetic mapping were as follows:

DtkMap1	5' TGGATGGCAGTAAGGGAGG 3'
	5' CTTAAGAGGGGCAAACCTGG 3'
DtkMap2	5' GCTTAGAGGAGGTGAGCCAGA 3'
	5' TGGGCAGTGCTGAGTTCC 3'

PCR was performed using standard conditions with the addition of ^{32}P-labelled dCTP. Specifically, 25 μl reactions were performed in 10 mM-Tris-HCl, 50 mM KCl using 250 ng of genomic DNA, 1 μM of each primer, and 1.4 mM $MgCl_2$. This was overlaid with oil, denatured at 94°C for 5 minutes, and transferred to an 80°C heating block. dNTPs were added to a final concentration of 0.2 mM, including 1.25 μCi of [α-^{32}P]dCTP (1μl of a 3000 Ci/mmole stock to 8 reactions). 1.25 units of AmpliTaq DNA polymerase (Perkin-Elmer Cetus) was added and cycling conditions were as follows: 58°C annealing reaction for 1 minute, 72°C extension reaction for 2 minutes, and 91°C denaturation for 1 minute. The cycle was repeated 30 times with a final 72°C extension reaction for 5 minutes. SSCP analysis was performed by electrophoresing the single-stranded PCR products on a non-denaturing gel as follows: 2 μl of the PCR reaction was added to 8 μl of USB stop solution (100% formamide containing xylene cyanol and bromophenol blue).

This was denatured for 5 minutes at 94°C and transferred to an ice bucket. 3μl was loaded on a 5% non-denaturing acrylamide gel containing 0.5X TBE and no glycerol. This was run in a 4°C cold room in 0.5X TBE at 40 watts constant power for 2-3 hours. The gel was transferred to filter paper, dried, and autoradiographed overnight with an intensifying screen.

Results

The chromosomal localisation of the gene encoding murine Dtk has been established on chromosome 2 band F using fluorescent in situ hybridisation. This result has been confirmed using single strand conformation polymorphism analysis in the BXD recombinant inbred series.

The gene encoding human Dtk has been mapped using fluorescent in situ hybridisation to chromosome 15q15.

C. Recombinant Expression of Dtks of the Invention

In yet another aspect, the present invention relates to the recombinant expression of the Dtks or of their extracellular receptor domains.

As will be exemplified below, the nucleic acid molecules that encode the receptors or the extracellular receptor domains of the invention may be inserted into known vectors for use in standard recombinant DNA techniques. Standard recombinant DNA techniques are those such as are described in Sambrook et al.; "Molecular Cloning" 2nd Edition Cold Spring Harbour Laboratory Press (1987) and by

Ausubel et al., Eds, "Current Protocols in Molecular Biology" Greene Publishing Associates and Wiley-Interscience, New York (1987).

Vectors for expressing proteins in bacteria, especially E. coli, are known. Such vectors include the PATH vectors described by Dieckmann and Tzagoloff in J. Biol. Chem. 260, 1513-1520 (1985). These vectors contain DNA sequences that encode anthranilate synthetase (TrpE) followed by a polylinker at the carboxy terminus. Other expression vector systems are based on beta-galactosidase (pGEX); lambda P maltose binding protein (pMAL); and gluthathione S-transferase (pGST) - see Gene 67, 31 (1988) and Peptide Research 3, 167 (1990).

Vectors useful in yeast are available and well known. A suitable example is the 2μ plasmid.

Suitable vectors for use in mammalian cells are also known. Such vectors include well-known derivatives of SV-40, adenovirus, retrovirus-derived DNA sequences and vectors derived from combination of plasmids and phage DNA.

Further eucaryotic expression vectors are known in the art (e.g. P.J. Southern and P. Berg, J. Mol. Appl. Genet. 1, 327-341 (1982); S. Subramani et al, Mol. Cell. Biol. 1, 854-864 (1981); R.J. Kaufmann and P.A. Sharp, "Amplification And Expression of Sequences Cotransfected with a Modular Dihydrofolate Reductase Complementary DNA Gene," J. Mol. Biol. 159, 601-621 (1982); R.J. Kaufmann and P.A. Sharp, Mol. Cell. Biol. 159, 601-664 (1982); S.I. Scahill et al, "Expression And Characterization Of The Product Of A Human Immune Interferon DNA

Gene In Chinese Hamster Ovary Cells," Proc. Natl. Acad. Sci. USA 80, 4654-4659 (1983); G. Urlaub and L.A. Chasin, Proc. Natl. Acad. Sci. USA 77, 4216-4220, (1980).

The expression vectors useful in the present invention contain at least one expression control sequence that is operatively linked to the DNA sequence or fragment to be expressed. The control sequence is inserted in the vector in order to control and to regulate the expression of the cloned DNA sequence. Examples of useful expression control sequences are the lac system, the trp system, the tac system, the trc system, major operator and promoter regions of phage lambda, the control region of fd coat protein, the glycolytic promoters of yeast, e.g. the promoter for 3-phosphoglycerate kinase, the promoters of yeast acid phosphatase, e.g. Pho5, the promoters of the yeast alpha-mating factors, and promoters derived from polyoma, adenovirus, retrovirus, and simian virus, e.g. the early and late promoters or SV40, and other sequences known to control the expression of genes of prokaryotic and eucaryotic cells and their viruses or combinations thereof.

Vectors containing the receptor-encoding DNA and control signals are inserted into a host cell for expression of the receptor. Some useful expression host cells include well-known prokaryotic and eucaryotic cells. Some suitable prokaryotic hosts include, for example, E. coli, such as E. coli SG-936, E. coli HB 101, E. coli W3110, E. coli X1776, E. coli X2282, E. coli DHT, and E. coli MR01, Pseudomonas, Bacillus, such as Bacillus subtilis, and Streptomyces. Suitable eucaryotic cells include yeast and other fungi, insect, animal cells, such as COS cells and CHO cells, human cells and plant cells in tissue culture.

A specific although non-limiting example of this aspect of the invention is set out below. It will be appreciated that while the expression of murine Dtk is exemplified, the procedures disclosed are equally applicable to the expression of other Dtks, or to the expression of extracellular receptor domains of such Dtks.

C.1 Expression of cloned murine Dtk in heterologous cell lines

The coding region of murine Dtk was ligated in-frame into the commercially available expression vector pcDNA3 (InVitrogen) using standard molecular biology techniques. The pcDNA3-Dtk construct was electroporated into several heterologous cell lines to demonstrate expression of Dtk. Electroporation, drug selection and isolation of Dtk-expressing clones for each cell line followed standard techniques (in M. Kriegler, "Gene Transfer and Expression - A Laboratory Manual", Stockton Press, New York 1990).

The Dtk construct was expressed in the factor-dependent cell lines FDC-P1, BAF/3 and 32D, and in the NIH 3T3 cell line (all commercially available). The expression of Dtk in these cell lines has been ascertained at the level of RNA using standard techniques for the isolation of RNA and its detection using radiolabelled Dtk probes which are familiar to those experienced in the field (see Sambrook et al., "Molecular Cloning," Second Edition, supra vide).

D. Ligands

The invention also includes ligands that bind to the Dtks of the invention.

The ligand may be a protein such as a growth factor that occurs naturally in a mammal, preferably the same mammal that produces the corresponding Dtk. The growth factor may be isolated and purified, or be present on the surface of an isolated population of cells, such as stromal cells.

The ligand may also be a molecule that does not occur naturally in a mammal. For example, antibodies raised against the receptors of the invention or against anti-ligand antibodies mimic the shape of, and act as, ligands if they constitute the negative image of the receptor or anti-ligand antibody binding site.

Such antibodies may be polyclonal but are preferably monoclonal. Monoclonal antibodies may be produced by methods known in the art. These methods include the immunological method described by Kohler and Milstein in Nature 256, 495-497 (1975) and Campbell in "Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas" in Burdon et al. Eds, Laboratory Techniques in Biochemistry and Molecular Biology, Volume 13, Elsevier Science Publishers, Amsterdam (1985); as well as by the recombinant DNA method described by Huse et al. in Science 246, 1275-1281 (1989).

In yet another form, the ligand may also be a non-protein molecule that acts as a ligand when it binds to, or otherwise comes into contact with, the receptor.

In addition, ligands may be of two functional types. The first functional type of ligand is a molecule which binds to the receptor and stimulates it in performing its normal function (a "stimulant ligand"). The second functional type of ligand is a

molecule which binds to the receptor and inhibits or prevents it performing its normal function (an "antagonistic ligand").

Both types of ligand will find application in either therapeutic or prophylactic treatments as described below.

D.1 <u>Sources of Ligands</u>

The strategy for isolating a ligand for the Dtks of the invention is based on the assumption that the ligand will either be a soluble, secreted protein or alternatively it will be membrane-bound or associated.

To screen for soluble ligands, conditioned media from a range of tumor cell lines and tissues can be used. Such cell lines are readily available from the American Type Culture Collection (ATCC) Rockville, Maryland, USA. Conditioned media is generated from these cell lines using a variety of culture and induction protocols. The cell lines are grown using standard tissue culture techniques which are detailed by ATCC for each cell line. Conditioned medium from tissues is generated by growing minced tissue fragments in culture medium for a defined time period.

To screen for membrane-associated ligands a different approach is taken. Cell lines in which from tissues which are in close proximity to those cells or tissues which have been shown to express the Dtk receptor are used. This approach is based on the likelihood of close cell-to-cell contact between receptor-expressing cells and ligand-expressing cells. An example of this is in the testis where Sertoli

cells express the receptor, while germ cells are considered a likely source of membrane-bound ligand. A further example in the brain would be where one type of neuron expresses the receptor, while microglial cells or another non-neuronal brain cell are considered likely to express the ligand.

D.2 Ligand Screening Procedures

In illustrating the screening procedures, reference will be made to murine Dtk as representative of the Dtks of the invention. Equivalent procedures can of course be employed in screening protocols using other mammalian Dtks such as human Dtk or the extracellular receptor domains of such Dtks.

Two approaches are followed to screen for the ligand for murine Dtk. If the ligand is soluble, assays which utilise either growth responses or changes in tyrosine phosphorylation will be used. Alternatively, if the ligand is membrane-bound, ligand-expressing cells will be detected using a Dtk-tag protein system whereby the extracellular domain of Dtk is fused with sequence encoding part of the human immunoglobulin molecule, such as the Fc region or the μ chain. The tag can then be detected using reagents which bind to the tag, such as Protein A-alkaline phosphatase or Protein A-radioiodine^{125}.

D.3 Soluble ligand

To detect soluble ligand in the media conditioned by tumor cell lines or tissues, a range of concentrations of this media are added to one of the factor-dependent cell lines described above, that have been transfected with, and express the Dtk receptor. These cell lines are routinely maintained in interleukin-3 containing

tissue culture medium. By withdrawing this medium and adding sources of potential ligand for Dtk, a growth response will be sought that is mediated via the introduced Dtk receptor. This response can be detected using the uptake of radiolabelled thymidine and counting this uptake by liquid scintillation spectroscopy. These techniques are standard for those familiar in the art (see Kriegler (supra); and Crosier et al., Proc. Natl. Acad. Sci. USA 88: 7744-8 (1991)).

An alternative detection system for ligands contained in tumor cell line conditioned medium uses the Dtk-expressing NIH 3T3 cell line as an indicator system, in conjunction with monitoring alterations in tyrosine phosphorylation of the Dtk receptor. Conditioned medium that contains the ligand for Dtk will trigger activation of the receptor which in turn is reflected in the phosphorylation status of the receptor. The system uses standard techniques whereby the NIH 3T3 cells are incubated with conditioned medium, cell lysates produced which in turn are immunoprecipitated with an anti-murine Dtk polyclonal antibody, proteins are resolved on SDS-PAGE gels, followed by transfer to nitrocellulose filters and subsequent Western blotting with an anti-phosphotyrosine antibody and detection using enhanced chemiluminescence techniques. These techniques are standard protein biochemistry methods (see B. Sefton and T. Hunter (eds), "Methods in Enzymology," vol 200 and 201, 1990; and Amersham, Manufacturer's protocols for ECL techniques). The expected result with this technique would be that potential ligand-containing media would stimulate increased tyrosine phosphorylation, compared with background levels detected in these cells.

D.4 Membrane-bound ligand

Screening for membrane-bound or associated ligands for the Dtk receptor relies on the use of a Dtk-tag fusion protein detection system. The extracellular domain of the Dtk receptor is fused in-frame to the Fc region of human immunoglobulin (IgG) or to part of the human μ chain of IgM. This procedure follows that described by Goodwin et al., Cell 73: 447-456 (1993). The fusion protein is produced by transfecting the fused genes contained within the expression pEDδc vector into COS cells. The fusion protein is purified on Protein A-Sepharose columns (Pharmacia). The Dtk-tag fusion proteins are biotinylated using sulfosuccinimiddyl-6-biotinamido)-hexanoate (Pierce Chemicals) according to the manufacturer's procedures. Alternatively, FITC-conjugated Dtk-tag fusion protein is generated by conjugating the fusion protein to FITC using standard techniques (see Suda et al., Cell 75: 169-1178, 1993).

The Dtk-tag fusion protein is used to screen for the expression of bound Dtk protein on tumor cell lines using flow cytometric techniques. The techniques used for the labelling of cells and flow cytometric analysis follow those described by Mosley et al., Cell 59: 335-348 (1989). Tumor cells are labelled on ice with the biotinylated Dtk-fusion protein using avidin-FITC, or the FITC-labelled protein is used directly in FACS analysis. The screening procedure is aimed at detecting a cell line that produces a signal above background with the Dtk-tag fusion protein, compared with an unrelated receptor-tag fusion protein. Sequential FACS sorting of Dtk ligand-expressing cells is undertaken to generate a high Dtk ligand-expressing tumor cell subline which can be used for the generation of a cDNA

expression library (for an overview of this strategy see Wong in Genetic Engineering Vol. 12, ed by J K Setlow, 1990).

E Expression Cloning of the Dtk Ligand

E.1 Construction of an expression library

A random-primed expression library is constructed from poly(A)$^+$ mRNA isolated from the cell line or tissue demonstrated to give a positive signal in either the growth assay, phosphorylation assay or Dtk-tag fusion protein assay outlined above. The techniques used for construction of the expression library are standard procedures for those experienced in the field (see McMahon et al., EMBO J. 10, 2821-2832, 1991; and Kriegler (supra)).

E.2 Cloning of the murine or human Dtk ligand

The expression library constructed from the cell line or tissue is screened by transfecting pools of cDNAs into COS cells using standard techniques (see Sambrook et al., supra). Two approaches are used to detect positive pools, depending on whether there has been evidence for either a soluble form of ligand or a membrane-bound form of ligand.

Soluble forms: COS supernatants are screened in the detection systems outlined above for soluble ligand forms. COS cells are grown in 10 cm plates using standard tissue culture techniques.

Membrane-bound forms: COS cells are grown in LabTech (Nunc) chambers and positive pools are detected by using the binding of Dtk-tag fusion protein to the

COS cells, followed by detection with either a Protein A-horseradish peroxidase enzymatic reaction or Protein A-^{125}I binding and subsequent autoradiography.

Procedures for the breaking down of cDNA pools, subsequent sib selection and the isolation of single cDNA clones are outlined in Sambrook et al., (supra) and Wong (supra). Sequence analysis of single cDNAs follows standard techniques. Once a single cDNA clone is isolated this is transfected into COS cells or into CHO cells for large scale production of protein using standard procedures.

F <u>Application of Ligands for the Dtks of the Invention</u>

The types of ligand discussed above can be employed in two distinctive methods in accordance with this invention.

The first such method is a method of stimulating the proliferation, differentiation and/or survival of a cell expressing a Dtk of the invention. This stimulation, which can occur *in vivo* or *ex vivo*, involves contacting the cell with an effective amount of the ligand.

The ability of a ligand according to the invention to stimulate cells such as stem cells which express the Dtk of the invention has important therapeutic applications. Such applications include medically treating mammals, including humans, whose stem cells do not sufficiently undergo self-renewal. Examples of such medical problems which can be treated in this way include those that occur when defects in haematopoietic stem cells or their related growth factors depress the number of blood cells, leading to disorders such as aplastic anaemia. The

treatment of bone marrow damage resulting from cancer chemotherapy and radiation is another example of a medical problem that could be treated in this way.

The method can also be applied in stimulating the proliferation, differentiation and/or survival of mammalian fetal or adult neuronal cells or cells that form part of the central nervous system. Again, this has important therapeutic applications. Such applications include treating mammals, including humans, for inherited or degenerative disorders of the central nervous system. An additional application is the treatment of individuals with central nervous system trauma, for example, spinal cord trauma resulting from either crushing or asphyxia.

Yet a further therapeutic application for the ligands of the invention is in sports medicine, particularly in the treatment of muscle injuries. The Dtk of the invention is abundantly expressed on myoblast cells but not on mature muscle cells. Application of the ligand will stimulate myoblast cell proliferation and differentiation, leading to muscle repair.

In terms of *ex vivo* applications, the method has implications for gene therapy. In gene therapy genes are inserted into host cells (such as haematopoietic stem cells and myoblasts) and the expression of the gene regulated by either an endogenous or an exogenous promoter. However, it is often difficult to maintain growth and survival of these cells *ex vivo* while they are being manipulated for the insertion of foreign genes. Therefore, as the Dtk of the invention is expressed on haematopoietic stem cells and myoblasts, the ligand has a direct application in

stimulating the growth, proliferation or simple survival of their cells during the manipulative process.

The second distinct method of the invention is a method of inhibiting the function of the Dtk of the invention. This method, which would normally be applied *in vivo* for both prophylactic and therapeutic applications, involves contacting the receptor with a ligand which blocks or prevents stimulation of the receptor (an antagonist ligand).

In terms of prophylaxis, such a method has specific application to the Sertoli cells of the testis, which abundantly express the receptor. Due to the involvement of these Sertoli cells in male fertility, contacting the receptors with an antagonistic ligand has a potential application in the control of male fertility including in male contraception.

A potential therapeutic application of contacting cells expressing the Dtk of the invention with an antagonistic ligand is in anti-tumour therapy. This potential application arises from the growing understanding of the role sometimes played by RTKs in tumour formation.

G <u>Therapeutic Applications of Soluble Receptors</u>

The extracellular receptor domain of the invention as described above also have potential therapeutic applications. Such applications are in a method of treating a disease, syndrome or condition caused or mediated by an excess of a ligand of the invention (whether stimulant or antagonistic).

In this method, the extracellular receptor domain of the Dtk in a soluble form can be used as a molecular "sponge" or "sink" to remove the excess of the ligand or at least to block its activity.

H Functional Equivalents

The invention includes functional equivalents of the Dtks, receptor domains, nucleic acid molecules and ligands described above.

The Dtks, extracellular receptor domains and ligands are or include proteins. A protein is considered a functional equivalent of another protein for a specific function if the equivalent protein is immunologically cross-reactive with, and has the same function as, the original protein. The equivalent may, for example, be a fragment of the protein, or a substitution, addition or deletion mutant of the protein.

For example, it is possible to substitute amino acids in a sequence with equivalent amino acids using conventional techniques. Groups of amino acids known normally to be equivalent are:

(a) Ala(A) Ser(S) Thr(T) Pro(P) Gly(G);

(b) Asn(N) Asp(D) Glu(E) Gln(Q);

(c) His(H) Arg(R) Lys(K);

(d) Met(M) Leu(L) Ile(I) Val(V); and

(e) Phe(F) Tyr(Y) Trp(W).

Substitutions, additions and/or deletions in the receptors and ligands may be made as long as the resulting equivalent receptors and ligands are immunologically cross-reactive with, and have the same function as, the native receptors and ligands.

The equivalent receptors and ligands will normally have substantially the same amino acid sequence as the native receptors and ligands. An amino acid sequence that is substantially the same as another sequence, but that differs from the other sequence by means of one or more substitutions, additions and/or deletions is considered to be an equivalent sequence. Preferably, less than 25%, more preferably less than 10%, and most preferably less than 5% of the number of amino acid residues in the amino acid sequence of the native receptors and ligands are substituted for, added to, or deleted from.

Equivalent nucleic acid molecules include nucleic acid sequences that encode equivalent receptors and ligands as defined above. Equivalent nucleic acid molecules also include nucleic acid sequences that, due to the degeneracy of the nucleic acid code, differ from native nucleic acid sequences in ways that do not affect the corresponding amino acid sequences.

Those persons skilled in the art will of course appreciate that the above description is provided by way of example only and that the invention is limited only by the lawful scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(1) AUCKLAND UNISERVICES LIMITED, a duly incorporated New Zealand company c/- The University of Auckland, 58 Symonds Street, Auckland, New Zeaalnd.

(2) TITLE OF INVENTION: Developmental Tyrosine Kinases and their Ligands.

(3) NUMBER OF SEQUENCES: 12

(4) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: A J PARK & SON

(B) STREET: HUDDART PARKER BUILDING, POST OFFICE SQUARE

(C) CITY: P O BOX 949, WELLINGTON

(D) COUNTRY: NEW ZEALAND

(5) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: 3.5,DS,HD FLOPPY DISC

(B) COMPUTER: IBM PC COMPATIBLE

(C) OPERATION SYSTEM: MS-DOS

(D) SOFTWARE: WORD PERFECT 5.1 FOR DOS

(6) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE: 16-FEBRUARY 1994

(C) CLASSIFICATION

(7) ATTORNEY/AGENT INFORMATION:

(A) NAME: BENNETT, MICHAEL R.

(8) TELECOMMUNICATION INFORMATION

(A) TELEPHONE: (64 4) 473 8278

(B) TELEFAX: (64 4) 472 3358

(2) INFORMATION FOR SEQUENCE ID NO. 1:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 874 AMINO ACIDS

(B) TYPE: AMINO ACID

(C) TOPOLOGY: LINEAR

(2) MOLECULE TYPE: PROTEIN

(3) SEQUENCE DESCRIPTION: SEQ ID NO. 1:

```
Met Gly Trp Pro Gly Leu Arg Pro Leu Leu Leu Ala Gly    13
Leu Ala Ser Leu Leu Leu Pro Gly Ser Ala Ala Ala Gly    26
Leu Lys Leu Met Gly Ala Pro Val Lys Met Thr Val Ser    39
Gln Gly Gln Pro Val Lys Leu Asn Cys Ser Val Glu Gly    52
Met Glu Asp Pro Asp Ile His Trp Met Lys Asp Gly Thr    65
Val Val Gln Asn Ala Ser Gln Val Ser Ile Ser Ile Ser    78
```

Glu His Ser Trp Ile Gly Leu Leu Ser Leu Lys Ser Val 91
Glu Arg Ser Asp Ala Gly Leu Tyr Trp Cys Gln Val Lys 104
Asp Gly Glu Glu Thr Lys Ile Ser Gln Ser Val Trp Leu 117
Thr Val Glu Gly Val Pro Phe Phe Thr Val Glu Pro Lys 130
Asp Leu Ala Val Pro Pro Asn Ala Pro Phe Gln Leu Ser 143
Cys Glu Ala Val Gly Pro Pro Glu Pro Val Thr Ile Tyr 156
Trp Trp Arg Gly Leu Thr Lys Val Gly Gly Pro Ala Pro 169
Ser Pro Ser Val Leu Asn Val Thr Gly Val Thr Gln Arg 182
Thr Glu Phe Ser Cys Glu Ala Arg Asn Ile Lys Gly Leu 195
Ala Thr Ser Arg Pro Ala Ile Val Arg Leu Gln Ala Pro 208
Pro Ala Ala Pro Phe Asn Thr Thr Val Thr Thr Ile Ser 221
Ser Tyr Asn Ala Ser Val Ala Trp Val Pro Gly Ala Asp 234
Gly Leu Ala Leu Leu His Ser Cys Thr Val Gln Val Ala 247
His Ala Pro Gly Glu Trp Glu Ala Leu Ala Val Val Val 260
Pro Val Pro Pro Phe Thr Cys Leu Leu Arg Asn Leu Ala 273
Pro Ala Thr Asn Tyr Ser Leu Arg Val Arg Cys Ala Asn 286
Ala Leu Gly Pro Ser Pro Tyr Gly Asp Trp Val Pro Phe 299
Gln Thr Lys Gly Leu Ala Pro Ala Arg Ala Pro Gln Asn 312
Phe His Ala Ile Arg Thr Asp Ser Gly Leu Ile Leu Glu 325
Trp Glu Glu Val Ile Pro Glu Asp Pro Gly Glu Gly Pro 338
Leu Gly Pro Tyr Lys Leu Ser Trp Val Gln Glu Asn Gly 351
Thr Gln Asp Glu Leu Met Val Glu Gly Thr Arg Ala Asn 364
Leu Thr Asp Trp Asp Pro Gln Lys Asp Leu Ile Leu Arg 377
Val Cys Ala Ser Asn Ala Ile Gly Asp Gly Pro Trp Ser 390
Gln Pro Leu Val Val Ser Ser His Asp His Ala Gly Arg 403
Gln Gly Pro Pro His Ser Arg Thr Ser Trp Val Pro Val 416
Val Leu Gly Val Leu Thr Ala Leu Ile Thr Ala Ala Ala 429
Leu Ala Leu Ile Leu Leu Arg Lys Arg Arg Lys Glu Thr 442
Arg Phe Gly Gln Ala Phe Asp Ser Val Met Ala Arg Gly 455
Glu Pro Ala Val His Phe Arg Ala Ala Arg Ser Phe Asn 468
Arg Glu Arg Pro Glu Arg Ile Glu Ala Thr Leu Asp Ser 481
Leu Gly Ile Ser Asp Glu Leu Lys Glu Lys Leu Glu Asp 494
Val Leu Ile Pro Glu Gln Gln Phe Thr Leu Gly Arg Met 507
Leu Gly Lys Gly Glu Phe Gly Ser Val Arg Glu Ala Gln 520

Leu	Lys	Gln	Glu	Asp	Gly	Ser	Phe	Val	Lys	Val	Ala	Val	533
Lys	Met	Leu	Lys	Ala	Asp	Ile	Ile	Ala	Ser	Ser	Asp	Ile	546
Glu	Glu	Phe	Leu	Arg	Glu	Ala	Ala	Cys	Met	Lys	Glu	Phe	559
Asp	His	Pro	His	Val	Ala	Lys	Leu	Val	Gly	Val	Ser	Leu	572
Arg	Ser	Arg	Ala	Lys	Gly	Arg	Leu	Pro	Ile	Pro	Met	Val	585
Ile	Leu	Pro	Phe	Met	Lys	His	Gly	Asp	Leu	His	Ala	Phe	598
Leu	Leu	Ala	Ser	Arg	Ile	Gly	Glu	Asn	Pro	Phe	Asn	Leu	611
Pro	Leu	Gln	Thr	Leu	Val	Arg	Phe	Met	Val	Asp	Ile	Ala	624
Cys	Gly	Met	Glu	Tyr	Leu	Ser	Ser	Arg	Asn	Phe	Ile	His	637
Arg	Asp	Leu	Ala	Ala	Arg	Asn	Cys	Met	Leu	Ala	Glu	Asp	650
Met	Thr	Val	Cys	Val	Ala	Asp	Phe	Gly	Leu	Ser	Arg	Lys	663
Ile	Tyr	Ser	Gly	Asp	Tyr	Tyr	Arg	Gln	Gly	Cys	Ala	Ser	676
Lys	Leu	Pro	Val	Lys	Trp	Leu	Ala	Leu	Glu	Ser	Leu	Ala	689
Asp	Asn	Leu	Tyr	Thr	Val	His	Ser	Asp	Val	Trp	Ala	Phe	702
Gly	Val	Thr	Met	Trp	Glu	Ile	Met	Thr	Arg	Gly	Gln	Thr	715
Pro	Tyr	Ala	Gly	Ile	Glu	Asn	Ala	Glu	Ile	Tyr	Asn	Tyr	728
Leu	Ile	Gly	Gly	Asn	Arg	Leu	Lys	Gln	Pro	Pro	Glu	Cys	741
Met	Glu	Glu	Val	Tyr	Asp	Leu	Met	Tyr	Gln	Cys	Trp	Ser	754
Ala	Asp	Pro	Lys	Gln	Arg	Pro	Ser	Phe	Thr	Cys	Leu	Arg	767
Met	Glu	Leu	Glu	Asn	Ile	Leu	Gly	His	Leu	Ser	Val	Leu	780
Ser	Thr	Ser	Gln	Asp	Pro	Leu	Tyr	Ile	Asn	Ile	Glu	Arg	793
Ala	Glu	Gln	Pro	Thr	Glu	Ser	Gly	Ser	Pro	Glu	Leu	His	806
Cys	Gly	Glu	Arg	Ser	Ser	Ser	Glu	Ala	Gly	Asp	Gly	Ser	819
Gly	Val	Gly	Ala	Val	Gly	Gly	Ile	Pro	Ser	Asp	Ser	Arg	832
Tyr	Ile	Phe	Ser	Pro	Gly	Gly	Leu	Ser	Glu	Ser	Pro	Gly	845
Gln	Leu	Glu	Gln	Gln	Pro	Glu	Ser	Pro	Leu	Asn	Glu	Asn	858
Gln	Arg	Leu	Leu	Leu	Leu	Gln	Gln	Gly	Leu	Leu	Pro	His	871
Ser	Ser	Cys											874

(3) INFORMATION FOR SEQUENCE ID NO. 2:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 850 AMINO ACIDS

(B) TYPE: AMINO ACID

(C) TOPOLOGY: LINEAR

(2) MOLECULE TYPE: PROTEIN

(3) SEQUENCE DESCRIPTION: SEQ ID NO. 2:

											Ala	Gly	2
Leu	Lys	Leu	Met	Gly	Ala	Pro	Val	Lys	Met	Thr	Val	Ser	15
Gln	Gly	Gln	Pro	Val	Lys	Leu	Asn	Cys	Ser	Val	Glu	Gly	28
Met	Glu	Asp	Pro	Asp	Ile	His	Trp	Met	Lys	Asp	Gly	Thr	41
Val	Val	Gln	Asn	Ala	Ser	Gln	Val	Ser	Ile	Ser	Ile	Ser	54
Glu	His	Ser	Trp	Ile	Gly	Leu	Leu	Ser	Leu	Lys	Ser	Val	67
Glu	Arg	Ser	Asp	Ala	Gly	Leu	Tyr	Trp	Cys	Gln	Val	Lys	80
Asp	Gly	Glu	Glu	Thr	Lys	Ile	Ser	Gln	Ser	Val	Trp	Leu	93
Thr	Val	Glu	Gly	Val	Pro	Phe	Phe	Thr	Val	Glu	Pro	Lys	106
Asp	Leu	Ala	Val	Pro	Pro	Asn	Ala	Pro	Phe	Gln	Leu	Ser	119
Cys	Glu	Ala	Val	Gly	Pro	Pro	Glu	Pro	Val	Thr	Ile	Tyr	132
Trp	Trp	Arg	Gly	Leu	Thr	Lys	Val	Gly	Gly	Pro	Ala	Pro	145
Ser	Pro	Ser	Val	Leu	Asn	Val	Thr	Gly	Val	Thr	Gln	Arg	158
Thr	Glu	Phe	Ser	Cys	Glu	Ala	Arg	Asn	Ile	Lys	Gly	Leu	171
Ala	Thr	Ser	Arg	Pro	Ala	Ile	Val	Arg	Leu	Gln	Ala	Pro	184
Pro	Ala	Ala	Pro	Phe	Asn	Thr	Thr	Val	Thr	Thr	Ile	Ser	197
Ser	Tyr	Asn	Ala	Ser	Val	Ala	Trp	Val	Pro	Gly	Ala	Asp	210
Gly	Leu	Ala	Leu	Leu	His	Ser	Cys	Thr	Val	Gln	Val	Ala	223
His	Ala	Pro	Gly	Glu	Trp	Glu	Ala	Leu	Ala	Val	Val	Val	236
Pro	Val	Pro	Pro	Phe	Thr	Cys	Leu	Leu	Arg	Asn	Leu	Ala	249
Pro	Ala	Thr	Asn	Tyr	Ser	Leu	Arg	Val	Arg	Cys	Ala	Asn	262
Ala	Leu	Gly	Pro	Ser	Pro	Tyr	Gly	Asp	Trp	Val	Pro	Phe	275
Gln	Thr	Lys	Gly	Leu	Ala	Pro	Ala	Arg	Ala	Pro	Gln	Asn	288
Phe	His	Ala	Ile	Arg	Thr	Asp	Ser	Gly	Leu	Ile	Leu	Glu	301
Trp	Glu	Glu	Val	Ile	Pro	Glu	Asp	Pro	Gly	Glu	Gly	Pro	314
Leu	Gly	Pro	Tyr	Lys	Leu	Ser	Trp	Val	Gln	Glu	Asn	Gly	327
Thr	Gln	Asp	Glu	Leu	Met	Val	Glu	Gly	Thr	Arg	Ala	Asn	340
Leu	Thr	Asp	Trp	Asp	Pro	Gln	Lys	Asp	Leu	Ile	Leu	Arg	353
Val	Cys	Ala	Ser	Asn	Ala	Ile	Gly	Asp	Gly	Pro	Trp	Ser	366
Gln	Pro	Leu	Val	Val	Ser	Ser	His	Asp	His	Ala	Gly	Arg	379
Gln	Gly	Pro	Pro	His	Ser	Arg	Thr	Ser	Trp	Val	Pro	Val	392
Val	Leu	Gly	Val	Leu	Thr	Ala	Leu	Ile	Thr	Ala	Ala	Ala	405

Leu	Ala	Leu	Ile	Leu	Leu	Arg	Lys	Arg	Arg	Lys	Glu	Thr	418
Arg	Phe	Gly	Gln	Ala	Phe	Asp	Ser	Val	Met	Ala	Arg	Gly	431
Glu	Pro	Ala	Val	His	Phe	Arg	Ala	Ala	Arg	Ser	Phe	Asn	444
Arg	Glu	Arg	Pro	Glu	Arg	Ile	Glu	Ala	Thr	Leu	Asp	Ser	457
Leu	Gly	Ile	Ser	Asp	Glu	Leu	Lys	Glu	Lys	Leu	Glu	Asp	470
Val	Leu	Ile	Pro	Glu	Gln	Gln	Phe	Thr	Leu	Gly	Arg	Met	483
Leu	Gly	Lys	Gly	Glu	Phe	Gly	Ser	Val	Arg	Glu	Ala	Gln	496
Leu	Lys	Gln	Glu	Asp	Gly	Ser	Phe	Val	Lys	Val	Ala	Val	509
Lys	Met	Leu	Lys	Ala	Asp	Ile	Ile	Ala	Ser	Ser	Asp	Ile	522
Glu	Glu	Phe	Leu	Arg	Glu	Ala	Ala	Cys	Met	Lys	Glu	Phe	535
Asp	His	Pro	His	Val	Ala	Lys	Leu	Val	Gly	Val	Ser	Leu	548
Arg	Ser	Arg	Ala	Lys	Gly	Arg	Leu	Pro	Ile	Pro	Met	Val	561
Ile	Leu	Pro	Phe	Met	Lys	His	Gly	Asp	Leu	His	Ala	Phe	574
Leu	Leu	Ala	Ser	Arg	Ile	Gly	Glu	Asn	Pro	Phe	Asn	Leu	587
Pro	Leu	Gln	Thr	Leu	Val	Arg	Phe	Met	Val	Asp	Ile	Ala	600
Cys	Gly	Met	Glu	Tyr	Leu	Ser	Ser	Arg	Asn	Phe	Ile	His	613
Arg	Asp	Leu	Ala	Ala	Arg	Asn	Cys	Met	Leu	Ala	Glu	Asp	626
Met	Thr	Val	Cys	Val	Ala	Asp	Phe	Gly	Leu	Ser	Arg	Lys	639
Ile	Tyr	Ser	Gly	Asp	Tyr	Tyr	Arg	Gln	Gly	Cys	Ala	Ser	652
Lys	Leu	Pro	Val	Lys	Trp	Leu	Ala	Leu	Glu	Ser	Leu	Ala	665
Asp	Asn	Leu	Tyr	Thr	Val	His	Ser	Asp	Val	Trp	Ala	Phe	678
Gly	Val	Thr	Met	Trp	Glu	Ile	Met	Thr	Arg	Gly	Gln	Thr	691
Pro	Tyr	Ala	Gly	Ile	Glu	Asn	Ala	Glu	Ile	Tyr	Asn	Tyr	704
Leu	Ile	Gly	Gly	Asn	Arg	Leu	Lys	Gln	Pro	Pro	Glu	Cys	717
Met	Glu	Glu	Val	Tyr	Asp	Leu	Met	Tyr	Gln	Cys	Trp	Ser	730
Ala	Asp	Pro	Lys	Gln	Arg	Pro	Ser	Phe	Thr	Cys	Leu	Arg	743
Met	Glu	Leu	Glu	Asn	Ile	Leu	Gly	His	Leu	Ser	Val	Leu	756
Ser	Thr	Ser	Gln	Asp	Pro	Leu	Tyr	Ile	Asn	Ile	Glu	Arg	769
Ala	Glu	Gln	Pro	Thr	Glu	Ser	Gly	Ser	Pro	Glu	Leu	His	782
Cys	Gly	Glu	Arg	Ser	Ser	Ser	Glu	Ala	Gly	Asp	Gly	Ser	795
Gly	Val	Gly	Ala	Val	Gly	Gly	Ile	Pro	Ser	Asp	Ser	Arg	808
Tyr	Ile	Phe	Ser	Pro	Gly	Gly	Leu	Ser	Glu	Ser	Pro	Gly	821
Gln	Leu	Glu	Gln	Gln	Pro	Glu	Ser	Pro	Leu	Asn	Glu	Asn	834
Gln	Arg	Leu	Leu	Leu	Leu	Gln	Gln	Gly	Leu	Leu	Pro	His	847

Ser Ser Cys 850

(4) INFORMATION FOR SEQUENCE ID NO. 3:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 876 AMINO ACIDS

(B) TYPE: AMINO ACID

(C) TOPOLOGY: LINEAR

(2) MOLECULE TYPE: PROTEIN

(3) SEQUENCE DESCRIPTION: SEQ ID NO. 3:

Met	Gly	Arg	Pro	Gly	Leu	Pro	Pro	Leu	Pro	Leu	Pro	Pro	13
Pro	Pro	Arg	Leu	Gly	Leu	Leu	Leu	Ala	Glu	Ser	Ala	Ala	26
Ala	Gly	Leu	Lys	Leu	Met	Gly	Ala	Pro	Val	Lys	Leu	Thr	39
Val	Ser	Gln	Gly	Gln	Pro	Val	Lys	Leu	Asn	Cys	Ser	Val	52
Glu	Gly	Met	Glu	Glu	Pro	Asp	Ile	Gln	Trp	Val	Lys	Asp	65
Gly	Ala	Val	Val	Gln	Asn	Leu	Asp	Gln	Leu	Tyr	Ile	Pro	78
Val	Ser	Glu	Gln	His	Trp	Ile	Gly	Phe	Leu	Ser	Leu	Lys	91
Ser	Val	Glu	Arg	Ser	Asp	Ala	Gly	Arg	Tyr	Trp	Cys	Gln	104
Val	Glu	Asp	Gly	Gly	Glu	Thr	Glu	Ile	Ser	Gln	Pro	Val	117
Trp	Leu	Thr	Val	Glu	Gly	Val	Pro	Phe	Phe	Thr	Val	Glu	130
Pro	Lys	Asp	Leu	Ala	Val	Pro	Pro	Asn	Ala	Pro	Phe	Gln	143
Leu	Ser	Cys	Glu	Ala	Val	Gly	Pro	Pro	Glu	Pro	Val	Thr	156
Ile	Val	Trp	Trp	Arg	Gly	Thr	Thr	Lys	Ile	Gly	Gly	Pro	169
Ala	Pro	Ser	Pro	Ser	Val	Leu	Asn	Val	Thr	Gly	Val	Thr	182
Gln	Ser	Thr	Met	Phe	Ser	Cys	Glu	Ala	His	Asn	Leu	Lys	195
Gly	Leu	Ala	Ser	Ser	Arg	Thr	Ala	Thr	Val	His	Leu	Gln	208
Ala	Leu	Pro	Ala	Ala	Pro	Phe	Asn	Ile	Thr	Val	Thr	Lys	221
Leu	Ser	Ser	Ser	Asn	Ala	Ser	Val	Ala	Trp	Met	Pro	Gly	234
Ala	Asp	Gly	Arg	Ala	Leu	Leu	Gln	Ser	Cys	Thr	Val	Gln	247
Val	Thr	Gln	Ala	Pro	Gly	Gly	Trp	Glu	Val	Leu	Ala	Val	260
Val	Val	Pro	Val	Pro	Pro	Phe	Thr	Cys	Leu	Leu	Arg	Asp	273
Leu	Val	Pro	Ala	Thr	Asn	Tyr	Ser	Leu	Arg	Val	Arg	Cys	286
Ala	Asn	Ala	Leu	Gly	Pro	Ser	Pro	Tyr	Ala	Asp	Trp	Val	299
Pro	Phe	Gln	Thr	Lys	Gly	Leu	Ala	Pro	Ala	Ser	Ala	Pro	312
Gln	Asn	Leu	His	Ala	Ile	Arg	Thr	Asp	Ser	Gly	Leu	Ile	325

Leu	Glu	Trp	Glu	Glu	Val	Ile	Pro	Glu	Ala	Pro	Leu	Glu	338
Gly	Pro	Leu	Gly	Pro	Tyr	Lys	Leu	Ser	Trp	Val	Gln	Asp	351
Asn	Gly	Thr	Gln	Asp	Glu	Leu	Thr	Val	Glu	Gly	Thr	Arg	364
Ala	Asn	Leu	Thr	Gly	Trp	Asp	Pro	Gln	Lys	Asp	Leu	Ile	377
Val	Arg	Val	Cys	Val	Ser	Asn	Ala	Val	Gly	Cys	Gly	Pro	390
Trp	Ser	Gln	Pro	Leu	Val	Val	Ser	Ser	His	Asp	Arg	Ala	403
Gly	Gln	Gln	Gly	Pro	Pro	His	Ser	Arg	Thr	Ser	Trp	Val	416
Pro	Val	Val	Leu	Gly	Val	Leu	Thr	Ala	Leu	Val	Thr	Ala	429
Ala	Ala	Leu	Ala	Leu	Ile	Leu	Leu	Arg	Lys	Arg	Arg	Lys	442
Glu	Thr	Arg	Phe	Gly	Gln	Ala	Phe	Asp	Ser	Val	Met	Ala	455
Arg	Gly	Glu	Pro	Ala	Val	His	Phe	Arg	Ala	Ala	Arg	Ser	468
Phe	Asn	Arg	Glu	Arg	Pro	Glu	Arg	Ile	Glu	Ala	Thr	Leu	481
Asp	Ser	Leu	Gly	Ile	Ser	Asp	Glu	Leu	Lys	Glu	Lys	Leu	494
Glu	Asp	Val	Leu	Ile	Pro	Glu	Gln	Gln	Phe	Thr	Leu	Gly	507
Arg	Met	Leu	Gly	Lys	Gly	Glu	Phe	Gly	Ser	Val	Arg	Glu	520
Ala	Gln	Leu	Lys	Gln	Glu	Asp	Gly	Ser	Phe	Val	Lys	Val	533
Ala	Val	Lys	Met	Leu	Lys	Ala	Asp	Ile	Ile	Ala	Ser	Ser	546
Asp	Ile	Glu	Glu	Phe	Leu	Arg	Glu	Ala	Ala	Cys	Met	Lys	559
Glu	Phe	Asp	His	Pro	His	Val	Ala	Lys	Leu	Val	Gly	Val	572
Ser	Leu	Arg	Ser	Arg	Ala	Lys	Gly	Arg	Leu	Pro	Ile	Pro	585
Met	Val	Ile	Leu	Pro	Phe	Met	Lys	His	Gly	Asp	Leu	His	598
Ala	Phe	Leu	Leu	Ala	Ser	Arg	Ile	Gly	Glu	Asn	Pro	Phe	611
Asn	Leu	Pro	Leu	Gln	Thr	Leu	Ile	Arg	Phe	Met	Val	Asp	624
Ile	Ala	Cys	Gly	Met	Glu	Tyr	Leu	Ser	Ser	Arg	Asn	Phe	637
Ile	His	Arg	Asp	Leu	Ala	Ala	Arg	Asn	Cys	Met	Leu	Ala	650
Glu	Asp	Met	Thr	Val	Cys	Val	Ala	Asp	Phe	Gly	Leu	Ser	663
Arg	Lys	Ile	Tyr	Ser	Gly	Asp	Tyr	Tyr	Arg	Gln	Gly	Cys	676
Ala	Ser	Lys	Leu	Pro	Val	Lys	Trp	Leu	Ala	Leu	Glu	Ser	689
Leu	Ala	Asp	Asn	Leu	Tyr	Thr	Val	Gln	Ser	Asp	Val	Trp	702
Ala	Phe	Gly	Val	Thr	Met	Trp	Glu	Ile	Met	Thr	Arg	Gly	715
Gln	Thr	Pro	Tyr	Ala	Gly	Ile	Glu	Asn	Ala	Glu	Ile	Tyr	728
Asn	Tyr	Leu	Ile	Gly	Gly	Asn	Arg	Leu	Lys	Gln	Pro	Pro	741
Glu	Cys	Met	Glu	Asp	Val	Tyr	Asp	Leu	Met	Tyr	Gln	Cys	754
Trp	Ser	Ala	Asp	Pro	Lys	Gln	Arg	Pro	Ser	Phe	Thr	Cys	767

Leu	Arg	Met	Glu	Leu	Glu	Asn	Ile	Leu	Gly	Gln	Leu	Ser	780
Val	Leu	Ser	Ala	Ser	Gln	Asp	Pro	Leu	Tyr	Ile	Asn	Ile	793
Glu	Arg	Ala	Glu	Glu	Pro	Thr	Val	Gly	Gly	Ser	Leu	Glu	806
Leu	Pro	Gly	Arg	Asp	Gln	Pro	Tyr	Ser	Gly	Ala	Gly	Asp	819
Gly	Ser	Gly	Met	Gly	Ala	Val	Gly	Gly	Thr	Pro	Ser	Asp	832
Cys	Arg	Tyr	Ile	Leu	Thr	Pro	Gly	Gly	Leu	Ala	Glu	Gln	845
Pro	Gly	Gln	Ala	Glu	His	Gln	Pro	Glu	Ser	Pro	Leu	Asn	858
Glu	Thr	Gln	Arg	Leu	Leu	Leu	Leu	Gln	Gln	Gly	Leu	Leu	871
Pro	His	Ser	Ser	Cys									876

(5) INFORMATION FOR SEQUENCE ID NO. 4:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 850 AMINO ACIDS

(B) TYPE: AMINO ACID

(C) TOPOLOGY: LINEAR

(2) MOLECULE TYPE: PROTEIN

(3) SEQUENCE DESCRIPTION: SEQ ID NO. 4:

Ala	Gly	Leu	Lys	Leu	Met	Gly	Ala	Pro	Val	Lys	Leu	Thr	13
Val	Ser	Gln	Gly	Gln	Pro	Val	Lys	Leu	Asn	Cys	Ser	Val	26
Glu	Gly	Met	Glu	Glu	Pro	Asp	Ile	Gln	Trp	Val	Lys	Asp	39
Gly	Ala	Val	Val	Gln	Asn	Leu	Asp	Gln	Leu	Tyr	Ile	Pro	52
Val	Ser	Glu	Gln	His	Trp	Ile	Gly	Phe	Leu	Ser	Leu	Lys	65
Ser	Val	Glu	Arg	Ser	Asp	Ala	Gly	Arg	Tyr	Trp	Cys	Gln	78
Val	Glu	Asp	Gly	Gly	Glu	Thr	Glu	Ile	Ser	Gln	Pro	Val	91
Trp	Leu	Thr	Val	Glu	Gly	Val	Pro	Phe	Phe	Thr	Val	Glu	104
Pro	Lys	Asp	Leu	Ala	Val	Pro	Pro	Asn	Ala	Pro	Phe	Gln	117
Leu	Ser	Cys	Glu	Ala	Val	Gly	Pro	Pro	Glu	Pro	Val	Thr	130
Ile	Val	Trp	Trp	Arg	Gly	Thr	Thr	Lys	Ile	Gly	Gly	Pro	143
Ala	Pro	Ser	Pro	Ser	Val	Leu	Asn	Val	Thr	Gly	Val	Thr	156
Gln	Ser	Thr	Met	Phe	Ser	Cys	Glu	Ala	His	Asn	Leu	Lys	169
Gly	Leu	Ala	Ser	Ser	Arg	Thr	Ala	Thr	Val	His	Leu	Gln	182
Ala	Leu	Pro	Ala	Ala	Pro	Phe	Asn	Ile	Thr	Val	Thr	Lys	195
Leu	Ser	Ser	Ser	Asn	Ala	Ser	Val	Ala	Trp	Met	Pro	Gly	208
Ala	Asp	Gly	Arg	Ala	Leu	Leu	Gln	Ser	Cys	Thr	Val	Gln	221
Val	Thr	Gln	Ala	Pro	Gly	Gly	Trp	Glu	Val	Leu	Ala	Val	234

Val	Val	Pro	Val	Pro	Pro	Phe	Thr	Cys	Leu	Leu	Arg	Asp	247
Leu	Val	Pro	Ala	Thr	Asn	Tyr	Ser	Leu	Arg	Val	Arg	Cys	260
Ala	Asn	Ala	Leu	Gly	Pro	Ser	Pro	Tyr	Ala	Asp	Trp	Val	273
Pro	Phe	Gln	Thr	Lys	Gly	Leu	Ala	Pro	Ala	Ser	Ala	Pro	286
Gln	Asn	Leu	His	Ala	Ile	Arg	Thr	Asp	Ser	Gly	Leu	Ile	299
Leu	Glu	Trp	Glu	Glu	Val	Ile	Pro	Glu	Ala	Pro	Leu	Glu	312
Gly	Pro	Leu	Gly	Pro	Tyr	Lys	Leu	Ser	Trp	Val	Gln	Asp	325
Asn	Gly	Thr	Gln	Asp	Glu	Leu	Thr	Val	Glu	Gly	Thr	Arg	338
Ala	Asn	Leu	Thr	Gly	Trp	Asp	Pro	Gln	Lys	Asp	Leu	Ile	351
Val	Arg	Val	Cys	Val	Ser	Asn	Ala	Val	Gly	Cys	Gly	Pro	364
Trp	Ser	Gln	Pro	Leu	Val	Val	Ser	Ser	His	Asp	Arg	Ala	377
Gly	Gln	Gln	Gly	Pro	Pro	His	Ser	Arg	Thr	Ser	Trp	Val	390
Pro	Val	Val	Leu	Gly	Val	Leu	Thr	Ala	Leu	Val	Thr	Ala	403
Ala	Ala	Leu	Ala	Leu	Ile	Leu	Leu	Arg	Lys	Arg	Arg	Lys	416
Glu	Thr	Arg	Phe	Gly	Gln	Ala	Phe	Asp	Ser	Val	Met	Ala	429
Arg	Gly	Glu	Pro	Ala	Val	His	Phe	Arg	Ala	Ala	Arg	Ser	442
Phe	Asn	Arg	Glu	Arg	Pro	Glu	Arg	Ile	Glu	Ala	Thr	Leu	455
Asp	Ser	Leu	Gly	Ile	Ser	Asp	Glu	Leu	Lys	Glu	Lys	Leu	468
Glu	Asp	Val	Leu	Ile	Pro	Glu	Gln	Gln	Phe	Thr	Leu	Gly	481
Arg	Met	Leu	Gly	Lys	Gly	Glu	Phe	Gly	Ser	Val	Arg	Glu	494
Ala	Gln	Leu	Lys	Gln	Glu	Asp	Gly	Ser	Phe	Val	Lys	Val	507
Ala	Val	Lys	Met	Leu	Lys	Ala	Asp	Ile	Ile	Ala	Ser	Ser	520
Asp	Ile	Glu	Glu	Phe	Leu	Arg	Glu	Ala	Ala	Cys	Met	Lys	533
Glu	Phe	Asp	His	Pro	His	Val	Ala	Lys	Leu	Val	Gly	Val	546
Ser	Leu	Arg	Ser	Arg	Ala	Lys	Gly	Arg	Leu	Pro	Ile	Pro	559
Met	Val	Ile	Leu	Pro	Phe	Met	Lys	His	Gly	Asp	Leu	His	572
Ala	Phe	Leu	Leu	Ala	Ser	Arg	Ile	Gly	Glu	Asn	Pro	Phe	585
Asn	Leu	Pro	Leu	Gln	Thr	Leu	Ile	Arg	Phe	Met	Val	Asp	598
Ile	Ala	Cys	Gly	Met	Glu	Tyr	Leu	Ser	Ser	Arg	Asn	Phe	611
Ile	His	Arg	Asp	Leu	Ala	Ala	Arg	Asn	Cys	Met	Leu	Ala	624
Glu	Asp	Met	Thr	Val	Cys	Val	Ala	Asp	Phe	Gly	Leu	Ser	637
Arg	Lys	Ile	Tyr	Ser	Gly	Asp	Tyr	Tyr	Arg	Gln	Gly	Cys	650
Ala	Ser	Lys	Leu	Pro	Val	Lys	Trp	Leu	Ala	Leu	Glu	Ser	663
Leu	Ala	Asp	Asn	Leu	Tyr	Thr	Val	Gln	Ser	Asp	Val	Trp	676

Ala	Phe	Gly	Val	Thr	Met	Trp	Glu	Ile	Met	Thr	Arg	Gly	689
Gln	Thr	Pro	Tyr	Ala	Gly	Ile	Glu	Asn	Ala	Glu	Ile	Tyr	702
Asn	Tyr	Leu	Ile	Gly	Gly	Asn	Arg	Leu	Lys	Gln	Pro	Pro	715
Glu	Cys	Met	Glu	Asp	Val	Tyr	Asp	Leu	Met	Tyr	Gln	Cys	728
Trp	Ser	Ala	Asp	Pro	Lys	Gln	Arg	Pro	Ser	Phe	Thr	Cys	741
Leu	Arg	Met	Glu	Leu	Glu	Asn	Ile	Leu	Gly	Gln	Leu	Ser	754
Val	Leu	Ser	Ala	Ser	Gln	Asp	Pro	Leu	Tyr	Ile	Asn	Ile	767
Glu	Arg	Ala	Glu	Glu	Pro	Thr	Val	Gly	Gly	Ser	Leu	Glu	780
Leu	Pro	Gly	Arg	Asp	Gln	Pro	Tyr	Ser	Gly	Ala	Gly	Asp	793
Gly	Ser	Gly	Met	Gly	Ala	Val	Gly	Gly	Thr	Pro	Ser	Asp	806
Cys	Arg	Tyr	Ile	Leu	Thr	Pro	Gly	Gly	Leu	Ala	Glu	Gln	819
Pro	Gly	Gln	Ala	Glu	His	Gln	Pro	Glu	Ser	Pro	Leu	Asn	832
Glu	Thr	Gln	Arg	Leu	Leu	Leu	Leu	Gln	Gln	Gly	Leu	Leu	845
Pro	His	Ser	Ser	Cys									850

(6) INFORMATION FOR SEQUENCE ID NO. 5:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 386 AMINO ACIDS

(B) TYPE: AMINO ACID

(C) TOPOLOGY: LINEAR

(2) MOLECULE TYPE: PROTEIN

(3) SEQUENCE DESCRIPTION: SEQ ID NO. 5:

											Ala	Gly	2
Leu	Lys	Leu	Met	Gly	Ala	Pro	Val	Lys	Met	Thr	Val	Ser	15
Gln	Gly	Gln	Pro	Val	Lys	Leu	Asn	Cys	Ser	Val	Glu	Gly	28
Met	Glu	Asp	Pro	Asp	Ile	His	Trp	Met	Lys	Asp	Gly	Thr	41
Val	Val	Gln	Asn	Ala	Ser	Gln	Val	Ser	Ile	Ser	Ile	Ser	54
Glu	His	Ser	Trp	Ile	Gly	Leu	Leu	Ser	Leu	Lys	Ser	Val	67
Glu	Arg	Ser	Asp	Ala	Gly	Leu	Tyr	Trp	Cys	Gln	Val	Lys	80
Asp	Gly	Glu	Glu	Thr	Lys	Ile	Ser	Gln	Ser	Val	Trp	Leu	93
Thr	Val	Glu	Gly	Val	Pro	Phe	Phe	Thr	Val	Glu	Pro	Lys	106
Asp	Leu	Ala	Val	Pro	Pro	Asn	Ala	Pro	Phe	Gln	Leu	Ser	119
Cys	Glu	Ala	Val	Gly	Pro	Pro	Glu	Pro	Val	Thr	Ile	Tyr	132
Trp	Trp	Arg	Gly	Leu	Thr	Lys	Val	Gly	Gly	Pro	Ala	Pro	145
Ser	Pro	Ser	Val	Leu	Asn	Val	Thr	Gly	Val	Thr	Gln	Arg	158

Thr	Glu	Phe	Ser	Cys	Glu	Ala	Arg	Asn	Ile	Lys	Gly	Leu	171
Ala	Thr	Ser	Arg	Pro	Ala	Ile	Val	Arg	Leu	Gln	Ala	Pro	184
Pro	Ala	Ala	Pro	Phe	Asn	Thr	Thr	Val	Thr	Thr	Ile	Ser	197
Ser	Tyr	Asn	Ala	Ser	Val	Ala	Trp	Val	Pro	Gly	Ala	Asp	210
Gly	Leu	Ala	Leu	Leu	His	Ser	Cys	Thr	Val	Gln	Val	Ala	223
His	Ala	Pro	Gly	Glu	Trp	Glu	Ala	Leu	Ala	Val	Val	Val	236
Pro	Val	Pro	Pro	Phe	Thr	Cys	Leu	Leu	Arg	Asn	Leu	Ala	249
Pro	Ala	Thr	Asn	Tyr	Ser	Leu	Arg	Val	Arg	Cys	Ala	Asn	262
Ala	Leu	Gly	Pro	Ser	Pro	Tyr	Gly	Asp	Trp	Val	Pro	Phe	275
Gln	Thr	Lys	Gly	Leu	Ala	Pro	Ala	Arg	Ala	Pro	Gln	Asn	288
Phe	His	Ala	Ile	Arg	Thr	Asp	Ser	Gly	Leu	Ile	Leu	Glu	301
Trp	Glu	Glu	Val	Ile	Pro	Glu	Asp	Pro	Gly	Glu	Gly	Pro	314
Leu	Gly	Pro	Tyr	Lys	Leu	Ser	Trp	Val	Gln	Glu	Asn	Gly	327
Thr	Gln	Asp	Glu	Leu	Met	Val	Glu	Gly	Thr	Arg	Ala	Asn	340
Leu	Thr	Asp	Trp	Asp	Pro	Gln	Lys	Asp	Leu	Ile	Leu	Arg	353
Val	Cys	Ala	Ser	Asn	Ala	Ile	Gly	Asp	Gly	Pro	Trp	Ser	366
Gln	Pro	Leu	Val	Val	Ser	Ser	His	Asp	His	Ala	Gly	Arg	379
Gln	Gly	Pro	Pro	His	Ser	Arg							386

(7) INFORMATION FOR SEQUENCE ID NO. 6:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 386 AMINO ACIDS

(B) TYPE: AMINO ACID

(C) TOPOLOGY: LINEAR

(2) MOLECULE TYPE: PROTEIN

(3) SEQUENCE DESCRIPTION: SEQ ID NO. 6:

Ala	Gly	Leu	Lys	Leu	Met	Gly	Ala	Pro	Val	Lys	Leu	Thr	13
Val	Ser	Gln	Gly	Gln	Pro	Val	Lys	Leu	Asn	Cys	Ser	Val	26
Glu	Gly	Met	Glu	Glu	Pro	Asp	Ile	Gln	Trp	Val	Lys	Asp	39
Gly	Ala	Val	Val	Gln	Asn	Leu	Asp	Gln	Leu	Tyr	Ile	Pro	52
Val	Ser	Glu	Gln	His	Trp	Ile	Gly	Phe	Leu	Ser	Leu	Lys	65
Ser	Val	Glu	Arg	Ser	Asp	Ala	Gly	Arg	Tyr	Trp	Cys	Gln	78
Val	Glu	Asp	Gly	Gly	Glu	Thr	Glu	Ile	Ser	Gln	Pro	Val	91
Trp	Leu	Thr	Val	Glu	Gly	Val	Pro	Phe	Phe	Thr	Val	Glu	104
Pro	Lys	Asp	Leu	Ala	Val	Pro	Pro	Asn	Ala	Pro	Phe	Gln	117

```
Leu  Ser  Cys  Glu  Ala  Val  Gly  Pro  Pro  Glu  Pro  Val  Thr    130
Ile  Val  Trp  Trp  Arg  Gly  Thr  Thr  Lys  Ile  Gly  Gly  Pro    143
Ala  Pro  Ser  Pro  Ser  Val  Leu  Asn  Val  Thr  Gly  Val  Thr    156
Gln  Ser  Thr  Met  Phe  Ser  Cys  Glu  Ala  His  Asn  Leu  Lys    169
Gly  Leu  Ala  Ser  Ser  Arg  Thr  Ala  Thr  Val  His  Leu  Gln    182
Ala  Leu  Pro  Ala  Ala  Pro  Phe  Asn  Ile  Thr  Val  Thr  Lys    195
Leu  Ser  Ser  Ser  Asn  Ala  Ser  Val  Ala  Trp  Met  Pro  Gly    208
Ala  Asp  Gly  Arg  Ala  Leu  Leu  Gln  Ser  Cys  Thr  Val  Gln    221
Val  Thr  Gln  Ala  Pro  Gly  Gly  Trp  Glu  Val  Leu  Ala  Val    234
Val  Val  Pro  Val  Pro  Pro  Phe  Thr  Cys  Leu  Leu  Arg  Asp    247
Leu  Val  Pro  Ala  Thr  Asn  Tyr  Ser  Leu  Arg  Val  Arg  Cys    260
Ala  Asn  Ala  Leu  Gly  Pro  Ser  Pro  Tyr  Ala  Asp  Trp  Val    273
Pro  Phe  Gln  Thr  Lys  Gly  Leu  Ala  Pro  Ala  Ser  Ala  Pro    286
Gln  Asn  Leu  His  Ala  Ile  Arg  Thr  Asp  Ser  Gly  Leu  Ile    299
Leu  Glu  Trp  Glu  Glu  Val  Ile  Pro  Glu  Ala  Pro  Leu  Glu    312
Gly  Pro  Leu  Gly  Pro  Tyr  Lys  Leu  Ser  Trp  Val  Gln  Asp    325
Asn  Gly  Thr  Gln  Asp  Glu  Leu  Thr  Val  Glu  Gly  Thr  Arg    338
Ala  Asn  Leu  Thr  Gly  Trp  Asp  Pro  Gln  Lys  Asp  Leu  Ile    351
Val  Arg  Val  Cys  Val  Ser  Asn  Ala  Val  Gly  Cys  Gly  Pro    364
Trp  Ser  Gln  Pro  Leu  Val  Val  Ser  Ser  His  Asp  Arg  Ala    377
Gly  Gln  Gln  Gly  Pro  Pro  His  Ser  Arg                        386
```

(8) INFORMATION FOR SEQUENCE ID NO. 7:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3919 BASE PAIRS

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: SINGLE

(D) TOPOLOGY: LINEAR

(2) MOLECULE TYPE: cDNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO. 7:

```
GGCACGAGTGTGGAAGGAGCGCGGTGGCCCAGCCGCAGCCCCGGGACTCCTCGCTGCTG    60
ACGGCGGTGGCCGCGGCTCTAGGCGGCCGCGGTCCCGGACGCCCCGGCCGAGCGCCGCC   120
CCCCGCCCCTCCCGCGGGCCTCCCGCCCCTCCTCCGCCACCCTCCTCTCAGCGCTCGCGG  180
GCCGGGCCCGGCATGGTGCGGCGTCGCCGCCGATGGCGCTGAGGCGGAGC            230

Met  Gly  Trp  Pro  Gly  Leu  Arg  Pro  Leu  Leu  Leu  Ala  Gly
ATG  GGG  TGG  CCG  GGG  CTC  CGG  CCG  CTG  CTG  CTG  GCG  GGA    269
```

Leu	Ala	Ser	Leu	Leu	Leu	Pro	Gly	Ser	Ala	Ala	Ala	Gly	
CTG	GCT	TCT	CTG	CTG	CTC	CCC	GGG	TCT	GCG	GCC	GCA	GGC	308
Leu	Lys	Leu	Met	Gly	Ala	Pro	Val	Lys	Met	Thr	Val	Ser	
CTG	AAG	CTC	ATG	GGC	GCC	CCA	GTG	AAG	ATG	ACC	GTG	TCT	347
Gln	Gly	Gln	Pro	Val	Lys	Leu	Asn	Cys	Ser	Val	Glu	Gly	
CAG	GGG	CAG	CCA	GTG	AAG	CTC	AAC	TGC	AGC	GTG	GAG	GGG	386
Met	Glu	Asp	Pro	Asp	Ile	His	Trp	Met	Lys	Asp	Gly	Thr	
ATG	GAG	GAC	CCT	GAC	ATC	CAC	TGG	ATG	AAG	GAT	GGC	ACC	425
Val	Val	Gln	Asn	Ala	Ser	Gln	Val	Ser	Ile	Ser	Ile	Ser	
GTG	GTC	CAG	AAT	GCA	AGC	CAG	GTG	TCC	ATC	TCC	ATC	AGC	464
Glu	His	Ser	Trp	Ile	Gly	Leu	Leu	Ser	Leu	Lys	Ser	Val	
GAG	CAC	AGC	TGG	ATT	GGC	TTA	CTC	AGC	CTA	AAG	TCA	GTG	503
Glu	Arg	Ser	Asp	Ala	Gly	Leu	Tyr	Trp	Cys	Gln	Val	Lys	
GAG	CGG	TCT	GAT	GCT	GGC	CTG	TAC	TGG	TGC	CAG	GTG	AAG	542
Asp	Gly	Glu	Glu	Thr	Lys	Ile	Ser	Gln	Ser	Val	Trp	Leu	
GAT	GGG	GAG	GAA	ACC	AAG	ATC	TCT	CAG	TCA	GTA	TGG	CTC	581
Thr	Val	Glu	Gly	Val	Pro	Phe	Phe	Thr	Val	Glu	Pro	Lys	
ACT	GTC	GAA	GGT	GTG	CCA	TTC	TTC	ACA	GTG	GAA	CCA	AAA	620
Asp	Leu	Ala	Val	Pro	Pro	Asn	Ala	Pro	Phe	Gln	Leu	Ser	
GAT	CTG	GCG	GTG	CCA	CCC	AAT	GCC	CCT	TTT	CAG	CTG	TCT	659
Cys	Glu	Ala	Val	Gly	Pro	Pro	Glu	Pro	Val	Thr	Ile	Tyr	
TGT	GAG	GCT	GTG	GGT	CCT	CCA	GAA	CCC	GTA	ACC	ATT	TAC	698
Trp	Trp	Arg	Gly	Leu	Thr	Lys	Val	Gly	Gly	Pro	Ala	Pro	
TGG	TGG	AGA	GGA	CTC	ACT	AAA	GTT	GGG	GGA	CCT	GCT	CCC	737
Ser	Pro	Ser	Val	Leu	Asn	Val	Thr	Gly	Val	Thr	Gln	Arg	
TCT	CCC	TCT	GTT	TTA	AAT	GTG	ACA	GGA	GTG	ACC	CAG	CGC	776
Thr	Glu	Phe	Ser	Cys	Glu	Ala	Arg	Asn	Ile	Lys	Gly	Leu	
ACA	GAG	TTT	TCT	TGT	GAA	GCC	CGC	AAC	ATA	AAA	GGC	CTG	815
Ala	Thr	Ser	Arg	Pro	Ala	Ile	Val	Arg	Leu	Gln	Ala	Pro	
GCC	ACT	TCC	CGA	CCA	GCC	ATT	GTT	CGC	CTT	CAA	GCA	CCG	854
Pro	Ala	Ala	Pro	Phe	Asn	Thr	Thr	Val	Thr	Thr	Ile	Ser	
CCT	GCA	GCT	CCT	TTC	AAC	ACC	ACA	GTA	ACA	ACG	ATC	TCC	893
Ser	Tyr	Asn	Ala	Ser	Val	Ala	Trp	Val	Pro	Gly	Ala	Asp	
AGC	TAC	AAC	GCT	AGC	GTG	GCC	TGG	GTG	CCA	GGT	GCT	GAC	932
Gly	Leu	Ala	Leu	Leu	His	Ser	Cys	Thr	Val	Gln	Val	Ala	
GGC	CTA	GCT	CTG	CTG	CAT	TCC	TGT	ACT	GTA	CAG	GTG	GCA	971
His	Ala	Pro	Gly	Glu	Trp	Glu	Ala	Leu	Ala	Val	Val	Val	
CAC	GCC	CCA	GGA	GAA	TGG	GAG	GCC	CTT	GCT	GTT	GTG	GTT	1010
Pro	Val	Pro	Pro	Phe	Thr	Cys	Leu	Leu	Arg	Asn	Leu	Ala	
CCT	GTG	CCA	CCT	TTT	ACC	TGC	CTG	CTT	CGG	AAC	TTG	GCC	1049
Pro	Ala	Thr	Asn	Tyr	Ser	Leu	Arg	Val	Arg	Cys	Ala	Asn	
CCT	GCC	ACC	AAC	TAC	AGC	CTT	AGG	GTG	CGC	TGT	GCC	AAT	1088
Ala	Leu	Gly	Pro	Ser	Pro	Tyr	Gly	Asp	Trp	Val	Pro	Phe	
GCC	TTG	GGC	CCT	TCT	CCC	TAC	GGC	GAC	TGG	GTG	CCC	TTT	1127
Gln	Thr	Lys	Gly	Leu	Ala	Pro	Ala	Arg	Ala	Pro	Gln	Asn	
CAG	ACA	AAG	GGC	CTA	GCG	CCA	GCC	AGA	GCT	CCT	CAG	AAT	1166

Phe His Ala Ile Arg Thr Asp Ser Gly Leu Ile Leu Glu
TTC CAT GCC ATT CGT ACC GAC TCA GGC CTT ATC CTG GAA 1205

Trp Glu Glu Val Ile Pro Glu Asp Pro Gly Glu Gly Pro
TGG GAA GAA GTG ATT CCT GAG GAC CCT GGG GAA GGC CCC 1244

Leu Gly Pro Tyr Lys Leu Ser Trp Val Gln Glu Asn Gly
CTA GGA CCT TAT AAG CTG TCC TGG GTC CAA GAA AAT GGA 1283

Thr Gln Asp Glu Leu Met Val Glu Gly Thr Arg Ala Asn
ACC CAG GAT GAG CTG ATG GTG GAA GGG ACC AGG GCC AAT 1322

Leu Thr Asp Trp Asp Pro Gln Lys Asp Leu Ile Leu Arg
CTG ACC GAC TGG GAT CCC CAG AAG GAC CTG ATT TTG CGT 1361

Val Cys Ala Ser Asn Ala Ile Gly Asp Gly Pro Trp Ser
GTG TGT GCC TCC AAT GCA ATT GGT GAT GGG CCC TGG AGT 1400

Gln Pro Leu Val Val Ser Ser His Asp His Ala Gly Arg
CAG CCA CTG GTG GTG TCT TCT CAT GAC CAT GCA GGG AGG 1439

Gln Gly Pro Pro His Ser Arg Thr Ser Trp Val Pro Val
CAG GGC CCT CCC CAC AGC CGC ACA TCC TGG GTG CCT GTG 1478

Val Leu Gly Val Leu Thr Ala Leu Ile Thr Ala Ala Ala
GTC CTG GGC GTG CTC ACC GCC CTG ATC ACA GCT GCT GCC 1517

Leu Ala Leu Ile Leu Leu Arg Lys Arg Arg Lys Glu Thr
TTG GCC CTC ATC CTG CTT CGG AAG AGA CGC AAG GAG ACG 1556

Arg Phe Gly Gln Ala Phe Asp Ser Val Met Ala Arg Gly
CGT TTC GGG CAA GCC TTT GAC AGT GTC ATG GCC CGA GGG 1595

Glu Pro Ala Val His Phe Arg Ala Ala Arg Ser Phe Asn
GAG CCA GCT GTA CAC TTC CGG GCA GCC CGA TCT TTC AAT 1634

Arg Glu Arg Pro Glu Arg Ile Glu Ala Thr Leu Asp Ser
CGA GAA AGG CCT GAA CGC ATT GAG GCC ACA TTG GAT AGC 1673

Leu Gly Ile Ser Asp Glu Leu Lys Glu Lys Leu Glu Asp
CTG GGC ATC AGC GAT GAA TTG AAG GAA AAG CTG GAG GAT 1712

Val Leu Ile Pro Glu Gln Gln Phe Thr Leu Gly Arg Met
GTC CTC ATT CCA GAG CAG CAG TTC ACC CTC GGT CGG ATG 1751

Leu Gly Lys Gly Glu Phe Gly Ser Val Arg Glu Ala Gln
TTG GGC AAA GGA GAG TTT GGA TCA GTG CGG GAA GCC CAG 1790

Leu Lys Gln Glu Asp Gly Ser Phe Val Lys Val Ala Val
CTA AAG CAG GAA GAT GGC TCC TTC GTG AAA GTG GCA GTG 1829

Lys Met Leu Lys Ala Asp Ile Ile Ala Ser Ser Asp Ile
AAG ATG CTG AAA GCT GAC ATC ATT GCC TCA AGC GAC ATA 1868

Glu Glu Phe Leu Arg Glu Ala Ala Cys Met Lys Glu Phe
GAA GAG TTC CTC CGG GAA GCA GCT TGC ATG AAG GAG TTT 1907

Asp His Pro His Val Ala Lys Leu Val Gly Val Ser Leu
GAC CAT CCA CAC GTG GCC AAG CTT GTT GGG GTG AGC CTC 1946

Arg Ser Arg Ala Lys Gly Arg Leu Pro Ile Pro Met Val
CGG AGC AGG GCT AAA GGT CGT CTC CCC ATT CCC ATG GTC 1985

Ile Leu Pro Phe Met Lys His Gly Asp Leu His Ala Phe
ATC CTG CCC TTC ATG AAA CAT GGA GAC TTG CAC GCC TTT 2024

Leu Leu Ala Ser Arg Ile Gly Glu Asn Pro Phe Asn Leu
CTG CTC GCC TCC CGA ATC GGG GAG AAC CCT TTT AAC CTG 2063

Pro	Leu	Gln	Thr	Leu	Val	Arg	Phe	Met	Val	Asp	Ile	Ala	
CCC	CTG	CAG	ACC	CTG	GTC	CGG	TTC	ATG	GTG	GAC	ATT	GCC	2102
Cys	Gly	Met	Glu	Tyr	Leu	Ser	Ser	Arg	Asn	Phe	Ile	His	
TGT	GGC	ATG	GAG	TAC	CTG	AGC	TCC	CGG	AAC	TTC	ATC	CAC	2141
Arg	Asp	Leu	Ala	Ala	Arg	Asn	Cys	Met	Leu	Ala	Glu	Asp	
CGA	GAC	CTA	GCA	GCT	CGG	AAT	TGC	ATG	CTG	GCC	GAG	GAC	2180
Met	Thr	Val	Cys	Val	Ala	Asp	Phe	Gly	Leu	Ser	Arg	Lys	
ATG	ACA	GTG	TGT	GTG	GCT	GAT	TTT	GGA	CTC	TCT	CGG	AAA	2219
Ile	Tyr	Ser	Gly	Asp	Tyr	Tyr	Arg	Gln	Gly	Cys	Ala	Ser	
ATC	TAT	AGC	GGG	GAC	TAT	TAT	CGT	CAG	GGC	TGT	GCC	TCC	2258
Lys	Leu	Pro	Val	Lys	Trp	Leu	Ala	Leu	Glu	Ser	Leu	Ala	
AAA	TTG	CCC	GTC	AAG	TGG	CTG	GCC	CTG	GAG	AGC	TTG	GCT	2297
Asp	Asn	Leu	Tyr	Thr	Val	His	Ser	Asp	Val	Trp	Ala	Phe	
GAC	AAC	TTG	TAT	ACT	GTA	CAC	AGT	GAT	GTG	TGG	GCC	TTC	2336
Gly	Val	Thr	Met	Trp	Glu	Ile	Met	Thr	Arg	Gly	Gln	Thr	
GGG	GTG	ACC	ATG	TGG	GAG	ATC	ATG	ACT	CGT	GGG	CAG	ACG	2375
Pro	Tyr	Ala	Gly	Ile	Glu	Asn	Ala	Glu	Ile	Tyr	Asn	Tyr	
CCA	TAT	GCT	GGC	ATT	GAA	AAT	GCT	GAG	ATT	TAC	AAC	TAC	2414
Leu	Ile	Gly	Gly	Asn	Arg	Leu	Lys	Gln	Pro	Pro	Glu	Cys	
CTC	ATC	GGC	GGG	AAC	CGC	CTG	AAG	CAG	CCT	CCG	GAG	TGC	2453
Met	Glu	Glu	Val	Tyr	Asp	Leu	Met	Tyr	Gln	Cys	Trp	Ser	
ATG	GAG	GAA	GTG	TAT	GAT	CTC	ATG	TAC	CAG	TGC	TGG	AGC	2492
Ala	Asp	Pro	Lys	Gln	Arg	Pro	Ser	Phe	Thr	Cys	Leu	Arg	
GCC	GAC	CCC	AAG	CAG	CGC	CCA	AGC	TTC	ACG	TGT	CTG	CGA	2531
Met	Glu	Leu	Glu	Asn	Ile	Leu	Gly	His	Leu	Ser	Val	Leu	
ATG	GAA	CTG	GAG	AAC	ATT	CTG	GGC	CAC	CTG	TCT	GTG	CTG	2370
Ser	Thr	Ser	Gln	Asp	Pro	Leu	Tyr	Ile	Asn	Ile	Glu	Arg	
TCC	ACC	AGC	CAG	GAC	CCC	TTG	TAC	ATC	AAC	ATT	GAG	AGA	2609
Ala	Glu	Gln	Pro	Thr	Glu	Ser	Gly	Ser	Pro	Glu	Leu	His	
GCT	GAG	CAG	CCT	ACT	GAG	AGT	GGC	AGC	CCT	GAG	CTG	CAC	2648
Cys	Gly	Glu	Arg	Ser	Ser	Ser	Glu	Ala	Gly	Asp	Gly	Ser	
TGT	GGA	GAG	CGA	TCC	AGC	AGC	GAG	GCA	GGG	GAC	GGC	AGT	2687
Gly	Val	Gly	Ala	Val	Gly	Gly	Ile	Pro	Ser	Asp	Ser	Arg	
GGC	GTG	GGG	GCA	GTA	GGT	GGC	ATC	CCC	AGT	GAC	TCT	CGG	2726
Tyr	Ile	Phe	Ser	Pro	Gly	Gly	Leu	Ser	Glu	Ser	Pro	Gly	
TAC	ATC	TTC	AGC	CCC	GGA	GGG	CTA	TCC	GAG	TCA	CCA	GGG	2765
Gln	Leu	Glu	Gln	Gln	Pro	Glu	Ser	Pro	Leu	Asn	Glu	Asn	
CAG	CTG	GAG	CAG	CAG	CCA	GAA	AGC	CCC	CTC	AAT	GAG	AAC	2804
Gln	Arg	Leu	Leu	Leu	Leu	Gln	Gln	Gly	Leu	Leu	Pro	His	
CAG	AGG	CTG	TTG	TTG	CTG	CAG	CAA	GGG	CTA	CTG	CCT	CAC	2843
Ser	Ser	Cys											
AGT	AGC	TGT											2852

*

TAACCCTCAGGCAGAGGAAAGTTGGGGCCCCTGGCTCTGCTGACCACTGTGCTGCCTGAC 2912
TAGGCCCAGTCTGATCACAGCCCAGGCAGCAAGGTATGGAGGCTCCTGTGGTAGCCCTCC 2972
CAAGCTGTGCTGGCGCCTGGACGGACCAAATTGCCCAATCCCAGTTCTTCCTGCAGCCGC 3032
TCTGGCCAGCCTGGCATCAGTTCAGGCCTTGGCTTAGAGGAGGTGAGCCAGAGCTGGTTG 3092

```
CCTGAATGCAGGCAGCTGGCAGGAGGGGAGGGTGGCTATGTTTCCATGGGTACCATGGGT    3152
GTGGATGGCAGTAAGGGAGGGTAGCAACAGCCCTGTGGCCCCTACCCTCCTGGCTGAGC    3212
TGCTCCTACTTTAGTGCATGCTTGGAGCCGCCTGCAGCCTGGAACTCAGCACTGCCCACC    3272
ACACTTGGGCCGAAATGCCAGGTTTGCCCCTCTTAAGTCACAAAGAGATGTCCATGTATT    3332
GTTCCCTTTTAGGTGATGATTAGGAAGGGATTGGCACACTTGGGTCCCTAAGCCCTATGG    3392
CAGGAAATGGTGGGATATTCTCAGGTCTGAATCCTCATCATCTTCCTGATTCCCCACCCT    3452
GCAAAGGCCTGGAACTGGCTGTGGGGCTCTGAGGCATGCTGAAGGACAAAAGATTACAGA    3512
GATCCGACTTCAAAAGGCAGGGTCTGAGTCTGGCAGGTGGAGAGGTGCTAAGGGGCTGGC    3572
CCAGGAGTCAGGCATTTCAGGACCCCTCCAAGCTTCTACAGTCTGTCTGAGCATGCTACC    3632
AAGCCCCCAGATACCCCAAAACTAACAGAGGCAGTTTTGTCTGAGCCCAGCCCTCCCACA    3692
TGATGACCCTTAGGTCTACCCTCCTCTCTAAATGGACATCCTCGTTTGTCCCAAGTCTCC    3752
AGAGAGACTACTGATGGCTGATGTGGGTAAGAAAAGTTCCAGGAACCAGGGCTGGGGTGG    3812
AACCAGGGCTGGGGTCGAGGCAGGCTCTTGGGCAGGCTCTTGCTGTTAGGAACATTTCTA    3872
AGCTATTAAGTTGCTGTTTCAAAACAAATAAAATTGAAACATAAAGAn                 3919
```

(9) INFORMATION FOR SEQUENCE ID NO. 8:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2550 BASE PAIRS

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: SINGLE

(D) TOPOLOGY: LINEAR

(2) MOLECULE TYPE: cDNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO. 8:

```
                                                        Ala Gly
                                                        GCA GGC      6

Leu Lys Leu Met Gly Ala Pro Val Lys Met Thr Val Ser
CTG AAG CTC ATG GGC GCC CCA GTG AAG ATG ACC GTG TCT     45

Gln Gly Gln Pro Val Lys Leu Asn Cys Ser Val Glu Gly
CAG GGG CAG CCA GTG AAG CTC AAC TGC AGC GTG GAG GGG     84

Met Glu Asp Pro Asp Ile His Trp Met Lys Asp Gly Thr
ATG GAG GAC CCT GAC ATC CAC TGG ATG AAG GAT GGC ACC    123

Val Val Gln Asn Ala Ser Gln Val Ser Ile Ser Ile Ser
GTG GTC CAG AAT GCA AGC CAG GTG TCC ATC TCC ATC AGC    162

Glu His Ser Trp Ile Gly Leu Leu Ser Leu Lys Ser Val
GAG CAC AGC TGG ATT GGC TTA CTC AGC CTA AAG TCA GTG    201

Glu Arg Ser Asp Ala Gly Leu Tyr Trp Cys Gln Val Lys
GAG CGG TCT GAT GCT GGC CTG TAC TGG TGC CAG GTG AAG    240

Asp Gly Glu Glu Thr Lys Ile Ser Gln Ser Val Trp Leu
GAT GGG GAG GAA ACC AAG ATC TCT CAG TCA GTA TGG CTC    279

Thr Val Glu Gly Val Pro Phe Phe Thr Val Glu Pro Lys
ACT GTC GAA GGT GTG CCA TTC TTC ACA GTG GAA CCA AAA    318

Asp Leu Ala Val Pro Pro Asn Ala Pro Phe Gln Leu Ser
GAT CTG GCG GTG CCA CCC AAT GCC CCT TTT CAG CTG TCT    357

Cys Glu Ala Val Gly Pro Pro Glu Pro Val Thr Ile Tyr
TGT GAG GCT GTG GGT CCT CCA GAA CCC GTA ACC ATT TAC    396

Trp Trp Arg Gly Leu Thr Lys Val Gly Gly Pro Ala Pro
TGG TGG AGA GGA CTC ACT AAA GTT GGG GGA CCT GCT CCC    435
```

Ser	Pro	Ser	Val	Leu	Asn	Val	Thr	Gly	Val	Thr	Gln	Arg	
TCT	CCC	TCT	GTT	TTA	AAT	GTG	ACA	GGA	GTG	ACC	CAG	CGC	474
Thr	Glu	Phe	Ser	Cys	Glu	Ala	Arg	Asn	Ile	Lys	Gly	Leu	
ACA	GAG	TTT	TCT	TGT	GAA	GCC	CGC	AAC	ATA	AAA	GGC	CTG	513
Ala	Thr	Ser	Arg	Pro	Ala	Ile	Val	Arg	Leu	Gln	Ala	Pro	
GCC	ACT	TCC	CGA	CCA	GCC	ATT	GTT	CGC	CTT	CAA	GCA	CCG	552
Pro	Ala	Ala	Pro	Phe	Asn	Thr	Thr	Val	Thr	Thr	Ile	Ser	
CCT	GCA	GCT	CCT	TTC	AAC	ACC	ACA	GTA	ACA	ACG	ATC	TCC	591
Ser	Tyr	Asn	Ala	Ser	Val	Ala	Trp	Val	Pro	Gly	Ala	Asp	
AGC	TAC	AAC	GCT	AGC	GTG	GCC	TGG	GTG	CCA	GGT	GCT	GAC	630
Gly	Leu	Ala	Leu	Leu	His	Ser	Cys	Thr	Val	Gln	Val	Ala	
GGC	CTA	GCT	CTG	CTG	CAT	TCC	TGT	ACT	GTA	CAG	GTG	GCA	669
His	Ala	Pro	Gly	Glu	Trp	Glu	Ala	Leu	Ala	Val	Val	Val	
CAC	GCC	CCA	GGA	GAA	TGG	GAG	GCC	CTT	GCT	GTT	GTG	GTT	708
Pro	Val	Pro	Pro	Phe	Thr	Cys	Leu	Leu	Arg	Asn	Leu	Ala	
CCT	GTG	CCA	CCT	TTT	ACC	TGC	CTG	CTT	CGG	AAC	TTG	GCC	747
Pro	Ala	Thr	Asn	Tyr	Ser	Leu	Arg	Val	Arg	Cys	Ala	Asn	
CCT	GCC	ACC	AAC	TAC	AGC	CTT	AGG	GTG	CGC	TGT	GCC	AAT	786
Ala	Leu	Gly	Pro	Ser	Pro	Tyr	Gly	Asp	Trp	Val	Pro	Phe	
GCC	TTG	GGC	CCT	TCT	CCC	TAC	GGC	GAC	TGG	GTG	CCC	TTT	825
Gln	Thr	Lys	Gly	Leu	Ala	Pro	Ala	Arg	Ala	Pro	Gln	Asn	
CAG	ACA	AAG	GGC	CTA	GCG	CCA	GCC	AGA	GCT	CCT	CAG	AAT	864
Phe	His	Ala	Ile	Arg	Thr	Asp	Ser	Gly	Leu	Ile	Leu	Glu	
TTC	CAT	GCC	ATT	CGT	ACC	GAC	TCA	GGC	CTT	ATC	CTG	GAA	903
Trp	Glu	Glu	Val	Ile	Pro	Glu	Asp	Pro	Gly	Glu	Gly	Pro	
TGG	GAA	GAA	GTG	ATT	CCT	GAG	GAC	CCT	GGG	GAA	GGC	CCC	942
Leu	Gly	Pro	Tyr	Lys	Leu	Ser	Trp	Val	Gln	Glu	Asn	Gly	
CTA	GGA	CCT	TAT	AAG	CTG	TCC	TGG	GTC	CAA	GAA	AAT	GGA	981
Thr	Gln	Asp	Glu	Leu	Met	Val	Glu	Gly	Thr	Arg	Ala	Asn	
ACC	CAG	GAT	GAG	CTG	ATG	GTG	GAA	GGG	ACC	AGG	GCC	AAT	1020
Leu	Thr	Asp	Trp	Asp	Pro	Gln	Lys	Asp	Leu	Ile	Leu	Arg	
CTG	ACC	GAC	TGG	GAT	CCC	CAG	AAG	GAC	CTG	ATT	TTG	CGT	1059
Val	Cys	Ala	Ser	Asn	Ala	Ile	Gly	Asp	Gly	Pro	Trp	Ser	
GTG	TGT	GCC	TCC	AAT	GCA	ATT	GGT	GAT	GGG	CCC	TGG	AGT	1098
Gln	Pro	Leu	Val	Val	Ser	Ser	His	Asp	His	Ala	Gly	Arg	
CAG	CCA	CTG	GTG	GTG	TCT	TCT	CAT	GAC	CAT	GCA	GGG	AGG	1137
Gln	Gly	Pro	Pro	His	Ser	Arg	Thr	Ser	Trp	Val	Pro	Val	
CAG	GGC	CCT	CCC	CAC	AGC	CGC	ACA	TCC	TGG	GTG	CCT	GTG	1176
Val	Leu	Gly	Val	Leu	Thr	Ala	Leu	Ile	Thr	Ala	Ala	Ala	
GTC	CTG	GGC	GTG	CTC	ACC	GCC	CTG	ATC	ACA	GCT	GCT	GCC	1215
Leu	Ala	Leu	Ile	Leu	Leu	Arg	Lys	Arg	Arg	Lys	Glu	Thr	
TTG	GCC	CTC	ATC	CTG	CTT	CGG	AAG	AGA	CGC	AAG	GAG	ACG	1254
Arg	Phe	Gly	Gln	Ala	Phe	Asp	Ser	Val	Met	Ala	Arg	Gly	
CGT	TTC	GGG	CAA	GCC	TTT	GAC	AGT	GTC	ATG	GCC	CGA	GGG	1293
Glu	Pro	Ala	Val	His	Phe	Arg	Ala	Ala	Arg	Ser	Phe	Asn	
GAG	CCA	GCT	GTA	CAC	TTC	CGG	GCA	GCC	CGA	TCT	TTC	AAT	1332

```
Arg Glu Arg Pro Glu Arg Ile Glu Ala Thr Leu Asp Ser
CGA GAA AGG CCT GAA CGC ATT GAG GCC ACA TTG GAT AGC   1371

Leu Gly Ile Ser Asp Glu Leu Lys Glu Lys Leu Glu Asp
CTG GGC ATC AGC GAT GAA TTG AAG GAA AAG CTG GAG GAT   1410

Val Leu Ile Pro Glu Gln Gln Phe Thr Leu Gly Arg Met
GTC CTC ATT CCA GAG CAG CAG TTC ACC CTC GGT CGG ATG   1449

Leu Gly Lys Gly Glu Phe Gly Ser Val Arg Glu Ala Gln
TTG GGC AAA GGA GAG TTT GGA TCA GTG CGG GAA GCC CAG   1488

Leu Lys Gln Glu Asp Gly Ser Phe Val Lys Val Ala Val
CTA AAG CAG GAA GAT GGC TCC TTC GTG AAA GTG GCA GTG   1527

Lys Met Leu Lys Ala Asp Ile Ile Ala Ser Ser Asp Ile
AAG ATG CTG AAA GCT GAC ATC ATT GCC TCA AGC GAC ATA   1566

Glu Glu Phe Leu Arg Glu Ala Ala Cys Met Lys Glu Phe
GAA GAG TTC CTC CGG GAA GCA GCT TGC ATG AAG GAG TTT   1605

Asp His Pro His Val Ala Lys Leu Val Gly Val Ser Leu
GAC CAT CCA CAC GTG GCC AAG CTT GTT GGG GTG AGC CTC   1644

Arg Ser Arg Ala Lys Gly Arg Leu Pro Ile Pro Met Val
CGG AGC AGG GCT AAA GGT CGT CTC CCC ATT CCC ATG GTC   1683

Ile Leu Pro Phe Met Lys His Gly Asp Leu His Ala Phe
ATC CTG CCC TTC ATG AAA CAT GGA GAC TTG CAC GCC TTT   1722

Leu Leu Ala Ser Arg Ile Gly Glu Asn Pro Phe Asn Leu
CTG CTC GCC TCC CGA ATC GGG GAG AAC CCT TTT AAC CTG   1761

Pro Leu Gln Thr Leu Val Arg Phe Met Val Asp Ile Ala
CCC CTG CAG ACC CTG GTC CGG TTC ATG GTG GAC ATT GCC   1800

Cys Gly Met Glu Tyr Leu Ser Ser Arg Asn Phe Ile His
TGT GGC ATG GAG TAC CTG AGC TCC CGG AAC TTC ATC CAC   1839

Arg Asp Leu Ala Ala Arg Asn Cys Met Leu Ala Glu Asp
CGA GAC CTA GCA GCT CGG AAT TGC ATG CTG GCC GAG GAC   1878

Met Thr Val Cys Val Ala Asp Phe Gly Leu Ser Arg Lys
ATG ACA GTG TGT GTG GCT GAT TTT GGA CTC TCT CGG AAA   1917

Ile Tyr Ser Gly Asp Tyr Tyr Arg Gln Gly Cys Ala Ser
ATC TAT AGC GGG GAC TAT TAT CGT CAG GGC TGT GCC TCC   1956

Lys Leu Pro Val Lys Trp Leu Ala Leu Glu Ser Leu Ala
AAA TTG CCC GTC AAG TGG CTG GCC CTG GAG AGC TTG GCT   1995

Asp Asn Leu Tyr Thr Val His Ser Asp Val Trp Ala Phe
GAC AAC TTG TAT ACT GTA CAC AGT GAT GTG TGG GCC TTC   2034

Gly Val Thr Met Trp Glu Ile Met Thr Arg Gly Gln Thr
GGG GTG ACC ATG TGG GAG ATC ATG ACT CGT GGG CAG ACG   2073

Pro Tyr Ala Gly Ile Glu Asn Ala Glu Ile Tyr Asn Tyr
CCA TAT GCT GGC ATT GAA AAT GCT GAG ATT TAC AAC TAC   2112

Leu Ile Gly Gly Asn Arg Leu Lys Gln Pro Pro Glu Cys
CTC ATC GGC GGG AAC CGC CTG AAG CAG CCT CCG GAG TGC   2151

Met Glu Glu Val Tyr Asp Leu Met Tyr Gln Cys Trp Ser
ATG GAG GAA GTG TAT GAT CTC ATG TAC CAG TGC TGG AGC   2190

Ala Asp Pro Lys Gln Arg Pro Ser Phe Thr Cys Leu Arg
GCC GAC CCC AAG CAG CGC CCA AGC TTC ACG TGT CTG CGA   2229
```

```
Met Glu Leu Glu Asn Ile Leu Gly His Leu Ser Val Leu
ATG GAA CTG GAG AAC ATT CTG GGC CAC CTG TCT GTG CTG  2268

Ser Thr Ser Gln Asp Pro Leu Tyr Ile Asn Ile Glu Arg
TCC ACC AGC CAG GAC CCC TTG TAC ATC AAC ATT GAG AGA  2307

Ala Glu Gln Pro Thr Glu Ser Gly Ser Pro Glu Leu His
GCT GAG CAG CCT ACT GAG AGT GGC AGC CCT GAG CTG CAC  2346

Cys Gly Glu Arg Ser Ser Ser Glu Ala Gly Asp Gly Ser
TGT GGA GAG CGA TCC AGC AGC GAG GCA GGG GAC GGC AGT  2385

Gly Val Gly Ala Val Gly Gly Ile Pro Ser Asp Ser Arg
GGC GTG GGG GCA GTA GGT GGC ATC CCC AGT GAC TCT CGG  2424

Tyr Ile Phe Ser Pro Gly Gly Leu Ser Glu Ser Pro Gly
TAC ATC TTC AGC CCC GGA GGG CTA TCC GAG TCA CCA GGG  2463

Gln Leu Glu Gln Gln Pro Glu Ser Pro Leu Asn Glu Asn
CAG CTG GAG CAG CAG CCA GAA AGC CCC CTC AAT GAG AAC  2502

Gln Arg Leu Leu Leu Leu Gln Gln Gly Leu Leu Pro His
CAG AGG CTG TTG TTG CTG CAG CAA GGG CTA CTG CCT CAC  2541

Ser Ser Cys
AGT AGC TGT                                          2550
```

(10) INFORMATION FOR SEQUENCE ID NO. 9:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4364 BASE PAIRS

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: SINGLE

(D) TOPOLOGY: LINEAR

(2) MOLECULE TYPE: cDNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO. 9:

```
CATTAGATCTTTACATGAAAGTAAAATTTATAAGATTTCTAGAAAGTCAAAAGATGATAA  60
CTATTTCTTAGGATACTAAAAGCACTCACATTATAGAAAAAAAATCAGTTAACTATACTC  120
CACAAACATTAAAGGCTCCCTATAAAAAAACATTTTTAATAGGCAAGCCACAGAAAGGGC  180
AAATATTAATAGTTTGCAATACATATGTATGAAAAGGAATTGAATCTAGAATATTTAACA  240
AAGCTTTACAACTCAAAAAATACAAAGAAAATATTTTTCTTCCAATTGGCAAATTACTTA  300
AACAGAACCTTCACAAAAGAAGATAAGAATGTTTAATAAACATTTGAAGCCATAATAATG  360
ACATCATTAGCCATGATGGAAATGCAAATTTAAGTACCACTTCACATCCACAAGAAAAAG  420
ATAAAAATAAAAGGACTGAGCTCACCAAACATTGGTGAGGATGTGGTAATACTGAAATTC  480
TTGTACCGTGCTCCTGAGGGTATAACATATTACAGGATTTTTTTGAAAACTAGTGGTTCC  540
TTATAAACTTAATGCCCTGGCAACCTCACACCTATTTACTTAAGAATGAAAGGCCCCCGC  600
CCTCCTCCCTCCTCGCTCGCGGGCCGGGCCCGGCATGGTGCGGCGTCGCCGCCGATGGCG  660
CTGAGGCGGAGC                                                  672

Met Gly Arg Pro Gly Leu Pro Pro Leu Pro Leu Pro Pro
ATG GGG CGG CCG GGG CTC CCG CCG CTG CCG CTG CCG CCG  711

Pro Pro Arg Leu Gly Leu Leu Leu Ala Glu Ser Ala Ala
CCA CCG CGG CTC GGG CTG CTG CTG GCG GAG TCC GCC GCC  750

Ala Gly Leu Lys Leu Met Gly Ala Pro Val Lys Leu Thr
GCA GGT CTG AAG CTC ATG GGA GCC CCG GTG AAG CTG ACA  789

Val Ser Gln Gly Gln Pro Val Lys Leu Asn Cys Ser Val
GTG TCT CAG GGG CAG CCG GTG AAG CTC AAC TGC AGT GTG  828
```

Glu Gly Met Glu Glu Pro Asp Ile Gln Trp Val Lys Asp
GAG GGG ATG GAG GAG CCT GAC ATC CAG TGG GTG AAG GAT 867

Gly Ala Val Val Gln Asn Leu Asp Gln Leu Tyr Ile Pro
GGG GCT GTG GTC CAG AAC TTG GAC CAG TTG TAC ATC CCA 906

Val Ser Glu Gln His Trp Ile Gly Phe Leu Ser Leu Lys
GTC AGC GAG CAG CAC TGG ATC GGC TTC CTC AGC CTG AAG 945

Ser Val Glu Arg Ser Asp Ala Gly Arg Tyr Trp Cys Gln
TCA GTG GAG CGC TCT GAC GCC GGC CGG TAC TGG TGC CAG 984

Val Glu Asp Gly Gly Glu Thr Glu Ile Ser Gln Pro Val
GTG GAG GAT GGG GGT GAA ACC GAG ATC TCC CAG CCA GTG 1023

Trp Leu Thr Val Glu Gly Val Pro Phe Phe Thr Val Glu
TGG CTC ACG GTA GAA GGT GTG CCA TTT TTC ACA GTG GAG 1062

Pro Lys Asp Leu Ala Val Pro Pro Asn Ala Pro Phe Gln
CCA AAA GAT CTG GCA GTG CCA CCC AAT GCC CCT TTC CAA 1101

Leu Ser Cys Glu Ala Val Gly Pro Pro Glu Pro Val Thr
CTG TCT TGT GAG GCT GTG GGT CCC CCT GAA CCT GTT ACC 1140

Ile Val Trp Trp Arg Gly Thr Thr Lys Ile Gly Gly Pro
ATT GTC TGG TGG AGA GGA ACT ACG AAG ATC GGG GGA CCC 1179

Ala Pro Ser Pro Ser Val Leu Asn Val Thr Gly Val Thr
GCT CCC TCT CCA TCT GTT TTA AAT GTA ACA GGG GTG ACC 1218

Gln Ser Thr Met Phe Ser Cys Glu Ala His Asn Leu Lys
CAG AGC ACC ATG TTT TCC TGT GAA GCT CAC AAC CTA AAA 1257

Gly Leu Ala Ser Ser Arg Thr Ala Thr Val His Leu Gln
GGC CTG GCC TCT TCT CGC ACA GCC ACT GTT CAC CTT CAA 1296

Ala Leu Pro Ala Ala Pro Phe Asn Ile Thr Val Thr Lys
GCA CTG CCT GCA GCC CCC TTC AAC ATC ACC GTG ACA AAG 1335

Leu Ser Ser Ser Asn Ala Ser Val Ala Trp Met Pro Gly
CTT TCC AGC AGC AAC GCT AGT GTG GCC TGG ATG CCA GGT 1374

Ala Asp Gly Arg Ala Leu Leu Gln Ser Cys Thr Val Gln
GCT GAT GGC CGA GCT CTG CTA CAG TCC TGT ACA GTT CAG 1413

Val Thr Gln Ala Pro Gly Gly Trp Glu Val Leu Ala Val
GTG ACA CAG GCC CCA GGA GGC TGG GAA GTC CTG GCT GTT 1452

Val Val Pro Val Pro Pro Phe Thr Cys Leu Leu Arg Asp
GTG GTC CCT GTG CCC CCC TTT ACC TGC CTG CTC CGG GAC 1491

Leu Val Pro Ala Thr Asn Tyr Ser Leu Arg Val Arg Cys
CTG GTG CCT GCC ACC AAC TAC AGC CTC AGG GTG CGC TGT 1530

Ala Asn Ala Leu Gly Pro Ser Pro Tyr Ala Asp Trp Val
GCC AAT GCC TTG GGG CCC TCT CCC TAT GCT GAC TGG GTG 1569

Pro Phe Gln Thr Lys Gly Leu Ala Pro Ala Ser Ala Pro
CCC TTT CAG ACC AAG GGT CTA GCC CCA GCC AGC GCT CCC 1608

Gln Asn Leu His Ala Ile Arg Thr Asp Ser Gly Leu Ile
CAA AAC CTC CAT GCC ATC CGC ACA GAT TCA GGC CTC ATC 1647

Leu Glu Trp Glu Glu Val Ile Pro Glu Ala Pro Leu Glu
TTG GAG TGG GAA GAA GTG ATC CCC GAG GCC CCT TTG GAA 1686

Gly Pro Leu Gly Pro Tyr Lys Leu Ser Trp Val Gln Asp
GGC CCC CTG GGA CCC TAC AAA CTG TCC TGG GTT CAA GAC 1725

Asn	Gly	Thr	Gln	Asp	Glu	Leu	Thr	Val	Glu	Gly	Thr	Arg	
AAT	GGA	ACC	CAG	GAT	GAG	CTG	ACA	GTG	GAG	GGG	ACC	AGG	1764
Ala	Asn	Leu	Thr	Gly	Trp	Asp	Pro	Gln	Lys	Asp	Leu	Ile	
GCC	AAT	TTG	ACA	GGC	TGG	GAT	CCC	CAA	AAG	GAC	CTG	ATC	1803
Val	Arg	Val	Cys	Val	Ser	Asn	Ala	Val	Gly	Cys	Gly	Pro	
GTA	CGT	GTG	TGC	GTC	TCC	AAT	GCA	GTT	GGC	TGT	GGA	CCC	1842
Trp	Ser	Gln	Pro	Leu	Val	Val	Ser	Ser	His	Asp	Arg	Ala	
TGG	AGT	CAG	CCA	CTG	GTG	GTC	TCT	TCT	CAT	GAC	CGT	GCA	1881
Gly	Gln	Gln	Gly	Pro	Pro	His	Ser	Arg	Thr	Ser	Trp	Val	
GGC	CAG	CAG	GGC	CCT	CCT	CAC	AGC	CGC	ACA	TCC	TGG	GTA	1920
Pro	Val	Val	Leu	Gly	Val	Leu	Thr	Ala	Leu	Val	Thr	Ala	
CCT	GTG	GTC	CTT	GGT	GTG	CTA	ACG	GCC	CTG	GTG	ACG	GCT	1959
Ala	Ala	Leu	Ala	Leu	Ile	Leu	Leu	Arg	Lys	Arg	Arg	Lys	
GCT	GCC	CTG	GCC	CTC	ATC	CTG	CTT	CGA	AAG	AGA	CGG	AAA	1998
Glu	Thr	Arg	Phe	Gly	Gln	Ala	Phe	Asp	Ser	Val	Met	Ala	
GAG	ACG	CGG	TTT	GGG	CAA	GCC	TTT	GAC	AGT	GTC	ATG	GCC	2037
Arg	Gly	Glu	Pro	Ala	Val	His	Phe	Arg	Ala	Ala	Arg	Ser	
CGG	GGA	GAG	CCA	GCC	GTT	CAC	TTC	CGG	GCA	GCC	CGG	TCC	2076
Phe	Asn	Arg	Glu	Arg	Pro	Glu	Arg	Ile	Glu	Ala	Thr	Leu	
TTC	AAT	CGA	GAA	AGG	CCC	GAG	CGC	ATC	GAG	GCC	ACA	TTG	2115
Asp	Ser	Leu	Gly	Ile	Ser	Asp	Glu	Leu	Lys	Glu	Lys	Leu	
GAC	AGC	TTG	GGC	ATC	AGC	GAT	GAA	CTA	AAG	GAA	AAA	CTG	2154
Glu	Asp	Val	Leu	Ile	Pro	Glu	Gln	Gln	Phe	Thr	Leu	Gly	
GAG	GAT	GTG	CTC	ATC	CCA	GAG	CAG	CAG	TTC	ACC	CTG	GGC	2193
Arg	Met	Leu	Gly	Lys	Gly	Glu	Phe	Gly	Ser	Val	Arg	Glu	
CGG	ATG	TTG	GGC	AAA	GGA	GAG	TTT	GGT	TCA	GTG	CGG	GAG	2232
Ala	Gln	Leu	Lys	Gln	Glu	Asp	Gly	Ser	Phe	Val	Lys	Val	
GCC	CAG	CTG	AAG	CAA	GAG	GAT	GGC	TCC	TTT	GTG	AAA	GTG	2271
Ala	Val	Lys	Met	Leu	Lys	Ala	Asp	Ile	Ile	Ala	Ser	Ser	
GCT	GTG	AAG	ATG	CTG	AAA	GCT	GAC	ATC	ATT	GCC	TCA	AGC	2310
Asp	Ile	Glu	Glu	Phe	Leu	Arg	Glu	Ala	Ala	Cys	Met	Lys	
GAC	ATT	GAA	GAG	TTC	CTC	AGG	GAA	GCA	GCT	TGC	ATG	AAG	2349
Glu	Phe	Asp	His	Pro	His	Val	Ala	Lys	Leu	Val	Gly	Val	
GAG	TTT	GAC	CAT	CCA	CAC	GTG	GCC	AAA	CTT	GTT	GGG	GTA	2388
Ser	Leu	Arg	Ser	Arg	Ala	Lys	Gly	Arg	Leu	Pro	Ile	Pro	
AGC	CTC	CGG	AGC	AGG	GCT	AAA	GGC	CGT	CTC	CCC	ATC	CCC	2427
Met	Val	Ile	Leu	Pro	Phe	Met	Lys	His	Gly	Asp	Leu	His	
ATG	GTC	ATC	TTG	CCC	TTC	ATG	AAG	CAT	GGG	GAC	CTG	CAT	2466
Ala	Phe	Leu	Leu	Ala	Ser	Arg	Ile	Gly	Glu	Asn	Pro	Phe	
GCC	TTC	CTG	CTC	GCC	TCC	CGG	ATT	GGG	GAG	AAC	CCC	TTT	2505
Asn	Leu	Pro	Leu	Gln	Thr	Leu	Ile	Arg	Phe	Met	Val	Asp	
AAC	CTA	CCC	CTC	CAG	ACC	CTG	ATC	CGG	TTC	ATG	GTG	GAC	2544
Ile	Ala	Cys	Gly	Met	Glu	Tyr	Leu	Ser	Ser	Arg	Asn	Phe	
ATT	GCC	TGC	GGC	ATG	GAG	TAC	CTG	AGC	TCT	CGG	AAC	TTC	2583
Ile	His	Arg	Asp	Leu	Ala	Ala	Arg	Asn	Cys	Met	Leu	Ala	
ATC	CAC	CGA	GAC	CTG	GCT	GCT	CGG	AAT	TGC	ATG	CTG	GCA	2622

```
Glu Asp Met Thr Val Cys Val Ala Asp Phe Gly Leu Ser
GAG GAC ATG ACA GTG TGT GTG GCT GAC TTC GGA CTC TCC      2661

Arg Lys Ile Tyr Ser Gly Asp Tyr Tyr Arg Gln Gly Cys
CGG AAG ATC TAC AGT GGG GAC TAC TAT CGT CAA GGC TGT      2700

Ala Ser Lys Leu Pro Val Lys Trp Leu Ala Leu Glu Ser
GCC TCC AAA CTG CCT GTC AAG TGG CTG GCC CTG GAG AGC      2739

Leu Ala Asp Asn Leu Tyr Thr Val Gln Ser Asp Val Trp
CTG GCC GAC AAC CTG TAT ACT GTG CAG AGT GAC GTG TGG      2778

Ala Phe Gly Val Thr Met Trp Glu Ile Met Thr Arg Gly
GCG TTC GGG GTG ACC ATG TGG GAG ATC ATG ACA CGT GGG      2817

Gln Thr Pro Tyr Ala Gly Ile Glu Asn Ala Glu Ile Tyr
CAG ACG CCA TAT GCT GGC ATC GAA AAC GCT GAG ATT TAC      2856

Asn Tyr Leu Ile Gly Gly Asn Arg Leu Lys Gln Pro Pro
AAC TAC CTC ATT GGC GGG AAC CGC CTG AAA CAG CCT CCG      2895

Glu Cys Met Glu Asp Val Tyr Asp Leu Met Tyr Gln Cys
GAG TGT ATG GAG GAC GTG TAT GAT CTC ATG TAC CAG TGC      2934

Trp Ser Ala Asp Pro Lys Gln Arg Pro Ser Phe Thr Cys
TGG AGT GCT GAC CCC AAG CAG CGC CCG AGC TTT ACT TGT      2973

Leu Arg Met Glu Leu Glu Asn Ile Leu Gly Gln Leu Ser
CTG CGA ATG GAA CTG GAG AAC ATC TTG GGC CAG CTG TCT      3012

Val Leu Ser Ala Ser Gln Asp Pro Leu Tyr Ile Asn Ile
GTG CTA TCT GCC AGC CAG GAC CCC TTA TAC ATC AAC ATC      3051

Glu Arg Ala Glu Glu Pro Thr Val Gly Gly Ser Leu Glu
GAG AGA GCT GAG GAG CCC ACT GTG GGA GGC AGC CTG GAG      3090

Leu Pro Gly Arg Asp Gln Pro Tyr Ser Gly Ala Gly Asp
CTA CCT GGC AGG GAT CAG CCC TAC AGT GGG GCT GGG GAT      3129

Gly Ser Gly Met Gly Ala Val Gly Gly Thr Pro Ser Asp
GGC AGT GGC ATG GGG GCA GTG GGT GGC ACT CCC AGT GAC      3168

Cys Arg Tyr Ile Leu Thr Pro Gly Gly Leu Ala Glu Gln
TGT CGG TAC ATA CTC ACC CCC GGA GGG CTG GCT GAG CAG      3207

Pro Gly Gln Ala Glu His Gln Pro Glu Ser Pro Leu Asn
CCA GGG CAG GCA GAG CAC CAG CCA GAG AGT CCC CTC AAT      3246

Glu Thr Gln Arg Leu Leu Leu Leu Gln Gln Gly Leu Leu
GAG ACA CAG AGG CTT TTG CTG CTG CAG CAA GGG CTA CTG      3285

Pro His Ser Ser Cys
CCA CAC AGT AGC TGT                                      3300

*
TAGCCCACAGGCAGAGGGCATCGGGGCCATTTGGCCGGCTCTGGTGGCCACTGAGCTGGC      3360
TGACTAAGCCCCGTCTGACCCCAGCCCAGACAGCAAGGTGTGGAGGCTCCTGTGGTAGTC      3420
CTCCCAAGCTGTGCTGGGAAGCCCGGACTGACCAAATCACCCAATCCCAGTTCTTCCTGC      3480
AACCACTCTGTGGCCAGCCTGGCATCAGTTTAGGCCTTGGCTTGATGGAAGTGGGCCAGT      3540
CCTGGTTGTCTGAACCCAGGCAGCTGGCAGGAGTGGGGTGGTTATGTTTCCATGGTTACC      3600
ATGGGTGTGGATGGCAGTGTGGGGAGGGCAGGTCCAGCTCTGTGGGCCCTACCCTCCTGC      3660
TGAGCTGCCCCTGCTGCTTAAGTGCATGCATTGAGCTGCCTCCAGCCTGGTGGCCCAGCT      3720
ATTACCACACTTGGGGTTTAAATATCCAGGTGTGCCCCTCCAAGTCAGAAAGAGATGTCC      3780
TTGTAATATTCCCTTTTAGGTGAGGGTTGGTAAGGGGTTGGTATCTCAGGTCTGAATCTT      3840
CACCATCTTTCTGATTCCGCACCCTGCCTACGCCAGGAGAAGTTGAGGGGAGCATGCTTC      3900
CCTGCAGCTGACCGGGTCACACAAAGGCATGCTGGAGTACCCAGCCTATCAGGTGCCCCT      3960
CTTCCAAAGGCAGCGTGCCGAGCCAGCAAGAGGAAGGGGTGCTGTGAGGCTTGCCCAGGA      4020
```

GCAAGTGAGGCCGGAGAGGAGTTCAGGAACCCTTCTCCATACCCACAATCTGAGCACGCT 4080
ACCAAATCTCAAAATATCCTAAGACTAACAAAGGCAGCTGTGTCTGAGCCCAACCCTTCT 4140
AAACGGTGACCTTTAGTGCCAACTTCCCCTCTAACTGGACAGCCTCTTCTGTCCCAAGTC 4200
TCCAGAGAGAAATCAGGCCTGATGAGGGGGAATTCCTGGAACCTGGACCCCAGCCTTGGT 4260
GGGGGAGCCTCTGGAATGCATGGGGCGGGTCCTAGCTGTTAGGGACATTTCCAAGCTGTT 4320
AGTTGCTGTTTAAAATAGAAATAAAATTGAAGACTAAAGACCTA$_n$ 4364

(11) INFORMATION FOR SEQUENCE ID NO. 10:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2550 BASE PAIRS

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: SINGLE

(D) TOPOLOGY: LINEAR

(2) MOLECULE TYPE: cDNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO. 10:

Ala Gly Leu Lys Leu Met Gly Ala Pro Val Lys Leu Thr
GCA GGT CTG AAG CTC ATG GGA GCC CCG GTG AAG CTG ACA 39

Val Ser Gln Gly Gln Pro Val Lys Leu Asn Cys Ser Val
GTG TCT CAG GGG CAG CCG GTG AAG CTC AAC TGC AGT GTG 78

Glu Gly Met Glu Glu Pro Asp Ile Gln Trp Val Lys Asp
GAG GGG ATG GAG GAG CCT GAC ATC CAG TGG GTG AAG GAT 117

Gly Ala Val Val Gln Asn Leu Asp Gln Leu Tyr Ile Pro
GGG GCT GTG GTC CAG AAC TTG GAC CAG TTG TAC ATC CCA 156

Val Ser Glu Gln His Trp Ile Gly Phe Leu Ser Leu Lys
GTC AGC GAG CAG CAC TGG ATC GGC TTC CTC AGC CTG AAG 195

Ser Val Glu Arg Ser Asp Ala Gly Arg Tyr Trp Cys Gln
TCA GTG GAG CGC TCT GAC GCC GGC CGG TAC TGG TGC CAG 234

Val Glu Asp Gly Gly Glu Thr Glu Ile Ser Gln Pro Val
GTG GAG GAT GGG GGT GAA ACC GAG ATC TCC CAG CCA GTG 273

Trp Leu Thr Val Glu Gly Val Pro Phe Phe Thr Val Glu
TGG CTC ACG GTA GAA GGT GTG CCA TTT TTC ACA GTG GAG 312

Pro Lys Asp Leu Ala Val Pro Pro Asn Ala Pro Phe Gln
CCA AAA GAT CTG GCA GTG CCA CCC AAT GCC CCT TTC CAA 351

Leu Ser Cys Glu Ala Val Gly Pro Pro Glu Pro Val Thr
CTG TCT TGT GAG GCT GTG GGT CCC CCT GAA CCT GTT ACC 390

Ile Val Trp Trp Arg Gly Thr Thr Lys Ile Gly Gly Pro
ATT GTC TGG TGG AGA GGA ACT ACG AAG ATC GGG GGA CCC 429

Ala Pro Ser Pro Ser Val Leu Asn Val Thr Gly Val Thr
GCT CCC TCT CCA TCT GTT TTA AAT GTA ACA GGG GTG ACC 468

Gln Ser Thr Met Phe Ser Cys Glu Ala His Asn Leu Lys
CAG AGC ACC ATG TTT TCC TGT GAA GCT CAC AAC CTA AAA 507

Gly Leu Ala Ser Ser Arg Thr Ala Thr Val His Leu Gln
GGC CTG GCC TCT TCT CGC ACA GCC ACT GTT CAC CTT CAA 546

Ala	Leu	Pro	Ala	Ala	Pro	Phe	Asn	Ile	Thr	Val	Thr	Lys	
GCA	CTG	CCT	GCA	GCC	CCC	TTC	AAC	ATC	ACC	GTG	ACA	AAG	585
Leu	Ser	Ser	Ser	Asn	Ala	Ser	Val	Ala	Trp	Met	Pro	Gly	
CTT	TCC	AGC	AGC	AAC	GCT	AGT	GTG	GCC	TGG	ATG	CCA	GGT	624
Ala	Asp	Gly	Arg	Ala	Leu	Leu	Gln	Ser	Cys	Thr	Val	Gln	
GCT	GAT	GGC	CGA	GCT	CTG	CTA	CAG	TCC	TGT	ACA	GTT	CAG	663
Val	Thr	Gln	Ala	Pro	Gly	Gly	Trp	Glu	Val	Leu	Ala	Val	
GTG	ACA	CAG	GCC	CCA	GGA	GGC	TGG	GAA	GTC	CTG	GCT	GTT	702
Val	Val	Pro	Val	Pro	Pro	Phe	Thr	Cys	Leu	Leu	Arg	Asp	
GTG	GTC	CCT	GTG	CCC	CCC	TTT	ACC	TGC	CTG	CTC	CGG	GAC	741
Leu	Val	Pro	Ala	Thr	Asn	Tyr	Ser	Leu	Arg	Val	Arg	Cys	
CTG	GTG	CCT	GCC	ACC	AAC	TAC	AGC	CTC	AGG	GTG	CGC	TGT	780
Ala	Asn	Ala	Leu	Gly	Pro	Ser	Pro	Tyr	Ala	Asp	Trp	Val	
GCC	AAT	GCC	TTG	GGG	CCC	TCT	CCC	TAT	GCT	GAC	TGG	GTG	819
Pro	Phe	Gln	Thr	Lys	Gly	Leu	Ala	Pro	Ala	Ser	Ala	Pro	
CCC	TTT	CAG	ACC	AAG	GGT	CTA	GCC	CCA	GCC	AGC	GCT	CCC	858
Gln	Asn	Leu	His	Ala	Ile	Arg	Thr	Asp	Ser	Gly	Leu	Ile	
CAA	AAC	CTC	CAT	GCC	ATC	CGC	ACA	GAT	TCA	GGC	CTC	ATC	897
Leu	Glu	Trp	Glu	Glu	Val	Ile	Pro	Glu	Ala	Pro	Leu	Glu	
TTG	GAG	TGG	GAA	GAA	GTG	ATC	CCC	GAG	GCC	CCT	TTG	GAA	936
Gly	Pro	Leu	Gly	Pro	Tyr	Lys	Leu	Ser	Trp	Val	Gln	Asp	
GGC	CCC	CTG	GGA	CCC	TAC	AAA	CTG	TCC	TGG	GTT	CAA	GAC	975
Asn	Gly	Thr	Gln	Asp	Glu	Leu	Thr	Val	Glu	Gly	Thr	Arg	
AAT	GGA	ACC	CAG	GAT	GAG	CTG	ACA	GTG	GAG	GGG	ACC	AGG	1014
Ala	Asn	Leu	Thr	Gly	Trp	Asp	Pro	Gln	Lys	Asp	Leu	Ile	
GCC	AAT	TTG	ACA	GGC	TGG	GAT	CCC	CAA	AAG	GAC	CTG	ATC	1053
Val	Arg	Val	Cys	Val	Ser	Asn	Ala	Val	Gly	Cys	Gly	Pro	
GTA	CGT	GTG	TGC	GTC	TCC	AAT	GCA	GTT	GGC	TGT	GGA	CCC	1092
Trp	Ser	Gln	Pro	Leu	Val	Val	Ser	Ser	His	Asp	Arg	Ala	
TGG	AGT	CAG	CCA	CTG	GTG	GTC	TCT	TCT	CAT	GAC	CGT	GCA	1131
Gly	Gln	Gln	Gly	Pro	Pro	His	Ser	Arg	Thr	Ser	Trp	Val	
GGC	CAG	CAG	GGC	CCT	CCT	CAC	AGC	CGC	ACA	TCC	TGG	GTA	1170
Pro	Val	Val	Leu	Gly	Val	Leu	Thr	Ala	Leu	Val	Thr	Ala	
CCT	GTG	GTC	CTT	GGT	GTG	CTA	ACG	GCC	CTG	GTG	ACG	GCT	1209
Ala	Ala	Leu	Ala	Leu	Ile	Leu	Leu	Arg	Lys	Arg	Arg	Lys	
GCT	GCC	CTG	GCC	CTC	ATC	CTG	CTT	CGA	AAG	AGA	CGG	AAA	1248
Glu	Thr	Arg	Phe	Gly	Gln	Ala	Phe	Asp	Ser	Val	Met	Ala	
GAG	ACG	CGG	TTT	GGG	CAA	GCC	TTT	GAC	AGT	GTC	ATG	GCC	1287
Arg	Gly	Glu	Pro	Ala	Val	His	Phe	Arg	Ala	Ala	Arg	Ser	
CGG	GGA	GAG	CCA	GCC	GTT	CAC	TTC	CGG	GCA	GCC	CGG	TCC	1326
Phe	Asn	Arg	Glu	Arg	Pro	Glu	Arg	Ile	Glu	Ala	Thr	Leu	
TTC	AAT	CGA	GAA	AGG	CCC	GAG	CGC	ATC	GAG	GCC	ACA	TTG	1365
Asp	Ser	Leu	Gly	Ile	Ser	Asp	Glu	Leu	Lys	Glu	Lys	Leu	
GAC	AGC	TTG	GGC	ATC	AGC	GAT	GAA	CTA	AAG	GAA	AAA	CTG	1404
Glu	Asp	Val	Leu	Ile	Pro	Glu	Gln	Gln	Phe	Thr	Leu	Gly	
GAG	GAT	GTG	CTC	ATC	CCA	GAG	CAG	CAG	TTC	ACC	CTG	GGC	1443

Arg	Met	Leu	Gly	Lys	Gly	Glu	Phe	Gly	Ser	Val	Arg	Glu	
CGG	ATG	TTG	GGC	AAA	GGA	GAG	TTT	GGT	TCA	GTG	CGG	GAG	1482
Ala	Gln	Leu	Lys	Gln	Glu	Asp	Gly	Ser	Phe	Val	Lys	Val	
GCC	CAG	CTG	AAG	CAA	GAG	GAT	GGC	TCC	TTT	GTG	AAA	GTG	1521
Ala	Val	Lys	Met	Leu	Lys	Ala	Asp	Ile	Ile	Ala	Ser	Ser	
GCT	GTG	AAG	ATG	CTG	AAA	GCT	GAC	ATC	ATT	GCC	TCA	AGC	1560
Asp	Ile	Glu	Glu	Phe	Leu	Arg	Glu	Ala	Ala	Cys	Met	Lys	
GAC	ATT	GAA	GAG	TTC	CTC	AGG	GAA	GCA	GCT	TGC	ATG	AAG	1599
Glu	Phe	Asp	His	Pro	His	Val	Ala	Lys	Leu	Val	Gly	Val	
GAG	TTT	GAC	CAT	CCA	CAC	GTG	GCC	AAA	CTT	GTT	GGG	GTA	1638
Ser	Leu	Arg	Ser	Arg	Ala	Lys	Gly	Arg	Leu	Pro	Ile	Pro	
AGC	CTC	CGG	AGC	AGG	GCT	AAA	GGC	CGT	CTC	CCC	ATC	CCC	1677
Met	Val	Ile	Leu	Pro	Phe	Met	Lys	His	Gly	Asp	Leu	His	
ATG	GTC	ATC	TTG	CCC	TTC	ATG	AAG	CAT	GGG	GAC	CTG	CAT	1716
Ala	Phe	Leu	Leu	Ala	Ser	Arg	Ile	Gly	Glu	Asn	Pro	Phe	
GCC	TTC	CTG	CTC	GCC	TCC	CGG	ATT	GGG	GAG	AAC	CCC	TTT	1755
Asn	Leu	Pro	Leu	Gln	Thr	Leu	Ile	Arg	Phe	Met	Val	Asp	
AAC	CTA	CCC	CTC	CAG	ACC	CTG	ATC	CGG	TTC	ATG	GTG	GAC	1794
Ile	Ala	Cys	Gly	Met	Glu	Tyr	Leu	Ser	Ser	Arg	Asn	Phe	
ATT	GCC	TGC	GGC	ATG	GAG	TAC	CTG	AGC	TCT	CGG	AAC	TTC	1833
Ile	His	Arg	Asp	Leu	Ala	Ala	Arg	Asn	Cys	Met	Leu	Ala	
ATC	CAC	CGA	GAC	CTG	GCT	GCT	CGG	AAT	TGC	ATG	CTG	GCA	1872
Glu	Asp	Met	Thr	Val	Cys	Val	Ala	Asp	Phe	Gly	Leu	Ser	
GAG	GAC	ATG	ACA	GTG	TGT	GTG	GCT	GAC	TTC	GGA	CTC	TCC	1911
Arg	Lys	Ile	Tyr	Ser	Gly	Asp	Tyr	Tyr	Arg	Gln	Gly	Cys	
CGG	AAG	ATC	TAC	AGT	GGG	GAC	TAC	TAT	CGT	CAA	GGC	TGT	1950
Ala	Ser	Lys	Leu	Pro	Val	Lys	Trp	Leu	Ala	Leu	Glu	Ser	
GCC	TCC	AAA	CTG	CCT	GTC	AAG	TGG	CTG	GCC	CTG	GAG	AGC	1989
Leu	Ala	Asp	Asn	Leu	Tyr	Thr	Val	Gln	Ser	Asp	Val	Trp	
CTG	GCC	GAC	AAC	CTG	TAT	ACT	GTG	CAG	AGT	GAC	GTG	TGG	2028
Ala	Phe	Gly	Val	Thr	Met	Trp	Glu	Ile	Met	Thr	Arg	Gly	
GCG	TTC	GGG	GTG	ACC	ATG	TGG	GAG	ATC	ATG	ACA	CGT	GGG	2067
Gln	Thr	Pro	Tyr	Ala	Gly	Ile	Glu	Asn	Ala	Glu	Ile	Tyr	
CAG	ACG	CCA	TAT	GCT	GGC	ATC	GAA	AAC	GCT	GAG	ATT	TAC	2106
Asn	Tyr	Leu	Ile	Gly	Gly	Asn	Arg	Leu	Lys	Gln	Pro	Pro	
AAC	TAC	CTC	ATT	GGC	GGG	AAC	CGC	CTG	AAA	CAG	CCT	CCG	2145
Glu	Cys	Met	Glu	Asp	Val	Tyr	Asp	Leu	Met	Tyr	Gln	Cys	
GAG	TGT	ATG	GAG	GAC	GTG	TAT	GAT	CTC	ATG	TAC	CAG	TGC	2184
Trp	Ser	Ala	Asp	Pro	Lys	Gln	Arg	Pro	Ser	Phe	Thr	Cys	
TGG	AGT	GCT	GAC	CCC	AAG	CAG	CGC	CCG	AGC	TTT	ACT	TGT	2223
Leu	Arg	Met	Glu	Leu	Glu	Asn	Ile	Leu	Gly	Gln	Leu	Ser	
CTG	CGA	ATG	GAA	CTG	GAG	AAC	ATC	TTG	GGC	CAG	CTG	TCT	2262
Val	Leu	Ser	Ala	Ser	Gln	Asp	Pro	Leu	Tyr	Ile	Asn	Ile	
GTG	CTA	TCT	GCC	AGC	CAG	GAC	CCC	TTA	TAC	ATC	AAC	ATC	2301
Glu	Arg	Ala	Glu	Glu	Pro	Thr	Val	Gly	Gly	Ser	Leu	Glu	
GAG	AGA	GCT	GAG	GAG	CCC	ACT	GTG	GGA	GGC	AGC	CTG	GAG	2340

Leu Pro Gly Arg Asp Gln Pro Tyr Ser Gly Ala Gly Asp
CTA CCT GGC AGG GAT CAG CCC TAC AGT GGG GCT GGG GAT 2379

Gly Ser Gly Met Gly Ala Val Gly Gly Thr Pro Ser Asp
GGC AGT GGC ATG GGG GCA GTG GGT GGC ACT CCC AGT GAC 2418

Cys Arg Tyr Ile Leu Thr Pro Gly Gly Leu Ala Glu Gln
TGT CGG TAC ATA CTC ACC CCC GGA GGG CTG GCT GAG CAG 2457

Pro Gly Gln Ala Glu His Gln Pro Glu Ser Pro Leu Asn
CCA GGG CAG GCA GAG CAC CAG CCA GAG AGT CCC CTC AAT 2496

Glu Thr Gln Arg Leu Leu Leu Leu Gln Gln Gly Leu Leu
GAG ACA CAG AGG CTT TTG CTG CTG CAG CAA GGG CTA CTG 2535

Pro His Ser Ser Cys
CCA CAC AGT AGC TGT 2550

(12) INFORMATION FOR SEQUENCE ID NO. 11:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1158 BASE PAIRS

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: SINGLE

(D) TOPOLOGY: LINEAR

(2) MOLECULE TYPE: cDNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO. 11:

Ala Gly
GCA GGC 6

Leu Lys Leu Met Gly Ala Pro Val Lys Met Thr Val Ser
CTG AAG CTC ATG GGC GCC CCA GTG AAG ATG ACC GTG TCT 45

Gln Gly Gln Pro Val Lys Leu Asn Cys Ser Val Glu Gly
CAG GGG CAG CCA GTG AAG CTC AAC TGC AGC GTG GAG GGG 84

Met Glu Asp Pro Asp Ile His Trp Met Lys Asp Gly Thr
ATG GAG GAC CCT GAC ATC CAC TGG ATG AAG GAT GGC ACC 123

Val Val Gln Asn Ala Ser Gln Val Ser Ile Ser Ile Ser
GTG GTC CAG AAT GCA AGC CAG GTG TCC ATC TCC ATC AGC 162

Glu His Ser Trp Ile Gly Leu Leu Ser Leu Lys Ser Val
GAG CAC AGC TGG ATT GGC TTA CTC AGC CTA AAG TCA GTG 201

Glu Arg Ser Asp Ala Gly Leu Tyr Trp Cys Gln Val Lys
GAG CGG TCT GAT GCT GGC CTG TAC TGG TGC CAG GTG AAG 240

Asp Gly Glu Glu Thr Lys Ile Ser Gln Ser Val Trp Leu
GAT GGG GAG GAA ACC AAG ATC TCT CAG TCA GTA TGG CTC 279

Thr Val Glu Gly Val Pro Phe Phe Thr Val Glu Pro Lys
ACT GTC GAA GGT GTG CCA TTC TTC ACA GTG GAA CCA AAA 318

Asp Leu Ala Val Pro Pro Asn Ala Pro Phe Gln Leu Ser
GAT CTG GCG GTG CCA CCC AAT GCC CCT TTT CAG CTG TCT 357

Cys Glu Ala Val Gly Pro Pro Glu Pro Val Thr Ile Tyr
TGT GAG GCT GTG GGT CCT CCA GAA CCC GTA ACC ATT TAC 396

```
Trp Trp Arg Gly Leu Thr Lys Val Gly Gly Pro Ala Pro
TGG TGG AGA GGA CTC ACT AAA GTT GGG GGA CCT GCT CCC      435

Ser Pro Ser Val Leu Asn Val Thr Gly Val Thr Gln Arg
TCT CCC TCT GTT TTA AAT GTG ACP GGA GTG ACC CAG CGC      474

Thr Glu Phe Ser Cys Glu Ala Arg Asn Ile Lys Gly Leu
ACA GAG TTT TCT TGT GAA GCC CGC AAC ATA AAA GGC CTG      513

Ala Thr Ser Arg Pro Ala Ile Val Arg Leu Gln Ala Pro
GCC ACT TCC CGA CCA GCC ATT GTT CGC CTT CAA GCA CCG      552

Pro Ala Ala Pro Phe Asn Thr Thr Val Thr Thr Ile Ser
CCT GCA GCT CCT TTC AAC ACC ACA GTA ACA ACG ATC TCC      591

Ser Tyr Asn Ala Ser Val Ala Trp Val Pro Gly Ala Asp
AGC TAC AAC GCT AGC GTG GCC TGG GTG CCA GGT GCT GAC      630

Gly Leu Ala Leu Leu His Ser Cys Thr Val Gln Val Ala
GGC CTA GCT CTG CTG CAT TCC TGT ACT GTA CAG GTG GCA      669

His Ala Pro Gly Glu Trp Glu Ala Leu Ala Val Val Val
CAC GCC CCA GGA GAA TGG GAG GCC CTT GCT GTT GTG GTT      708

Pro Val Pro Pro Phe Thr Cys Leu Leu Arg Asn Leu Ala
CCT GTG CCA CCT TTT ACC TGC CTG CTT CGG AAC TTG GCC      747

Pro Ala Thr Asn Tyr Ser Leu Arg Val Arg Cys Ala Asn
CCT GCC ACC AAC TAC AGC CTT AGG GTG CGC TGT GCC AAT      786

Ala Leu Gly Pro Ser Pro Tyr Gly Asp Trp Val Pro Phe
GCC TTG GGC CCT TCT CCC TAC GGC GAC TGG GTG CCC TTT      825

Gln Thr Lys Gly Leu Ala Pro Ala Arg Ala Pro Gln Asn
CAG ACA AAG GGC CTA GCG CCA GCC AGA GCT CCT CAG AAT      864

Phe His Ala Ile Arg Thr Asp Ser Gly Leu Ile Leu Glu
TTC CAT GCC ATT CGT ACC GAC TCA GGC CTT ATC CTG GAA      903

Trp Glu Glu Val Ile Pro Glu Asp Pro Gly Glu Gly Pro
TGG GAA GAA GTG ATT CCT GAG GAC CCT GGG GAA GGC CCC      941

Leu Gly Pro Tyr Lys Leu Ser Trp Val Gln Glu Asn Gly
CTA GGA CCT TAT AAG CTG TCC TGG GTC CAA GAA AAT GGA      981

Thr Gln Asp Glu Leu Met Val Glu Gly Thr Arg Ala Asn
ACC CAG GAT GAG CTG ATG GTG GAA GGG ACC AGG GCC AAT      1020

Leu Thr Asp Trp Asp Pro Gln Lys Asp Leu Ile Leu Arg
CTG ACC GAC TGG GAT CCC CAG AAG GAC CTG ATT TTG CGT      1059

Val Cys Ala Ser Asn Ala Ile Gly Asp Gly Pro Trp Ser
GTG TGT GCC TCC AAT GCA ATT GGT GAT GGG CCC TGG AGT      1098

Gln Pro Leu Val Val Ser Ser His Asp His Ala Gly Arg
CAG CCA CTG GTG GTG TCT TCT CAT GAC CAT GCA GGG AGG      1137

Gln Gly Pro Pro His Ser Arg
CAG GGC CCT CCC CAC AGC CGC                              1158
```

(13) INFORMATION FOR SEQUENCE ID NO. 12:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1158 BASE PAIRS

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: SINGLE

(D) TOPOLOGY: LINEAR

(2) MOLECULE TYPE: cDNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO. 12:

```
Ala Gly Leu Lys Leu Met Gly Ala Pro Val Lys Leu Thr
GCA GGT CTG AAG CTC ATG GGA GCC CCG GTG AAG CTG ACA      39

Val Ser Gln Gly Gln Pro Val Lys Leu Asn Cys Ser Val
GTG TCT CAG GGG CAG CCG GTG AAG CTC AAC TGC AGT GTG      78

Glu Gly Met Glu Glu Pro Asp Ile Gln Trp Val Lys Asp
GAG GGG ATG GAG GAG CCT GAC ATC CAG TGG GTG AAG GAT     117

Gly Ala Val Val Gln Asn Leu Asp Gln Leu Tyr Ile Pro
GGG GCT GTG GTC CAG AAC TTG GAC CAG TTG TAC ATC CCA     156

Val Ser Glu Gln His Trp Ile Gly Phe Leu Ser Leu Lys
GTC AGC GAG CAG CAC TGG ATC GGC TTC CTC AGC CTG AAG     195

Ser Val Glu Arg Ser Asp Ala Gly Arg Tyr Trp Cys Gln
TCA GTG GAG CGC TCT GAC GCC GGC CGG TAC TGG TGC CAG     234

Val Glu Asp Gly Gly Glu Thr Glu Ile Ser Gln Pro Val
GTG GAG GAT GGG GGT GAA ACC GAG ATC TCC CAG CCA GTG     273

Trp Leu Thr Val Glu Gly Val Pro Phe Phe Thr Val Glu
TGG CTC ACG GTA GAA GGT GTG CCA TTT TTC ACA GTG GAG     312

Pro Lys Asp Leu Ala Val Pro Pro Asn Ala Pro Phe Gln
CCA AAA GAT CTG GCA GTG CCA CCC AAT GCC CCT TTC CAA     351

Leu Ser Cys Glu Ala Val Gly Pro Pro Glu Pro Val Thr
CTG TCT TGT GAG GCT GTG GGT CCC CCT GAA CCT GTT ACC     390

Ile Val Trp Trp Arg Gly Thr Thr Lys Ile Gly Gly Pro
ATT GTC TGG TGG AGA GGA ACT ACG AAG ATC GGG GGA CCC     429

Ala Pro Ser Pro Ser Val Leu Asn Val Thr Gly Val Thr
GCT CCC TCT CCA TCT GTT TTA AAT GTA ACA GGG GTG ACC     468

Gln Ser Thr Met Phe Ser Cys Glu Ala His Asn Leu Lys
CAG AGC ACC ATG TTT TCC TGT GAA GCT CAC AAC CTA AAA     507

Gly Leu Ala Ser Ser Arg Thr Ala Thr Val His Leu Gln
GGC CTG GCC TCT TCT CGC ACA GCC ACT GTT CAC CTT CAA     546

Ala Leu Pro Ala Ala Pro Phe Asn Ile Thr Val Thr Lys
GCA CTG CCT GCA GCC CCC TTC AAC ATC ACC GTG ACA AAG     585

Leu Ser Ser Ser Asn Ala Ser Val Ala Trp Met Pro Gly
CTT TCC AGC AGC AAC GCT AGT GTG GCC TGG ATG CCA GGT     624

Ala Asp Gly Arg Ala Leu Leu Gln Ser Cys Thr Val Gln
GCT GAT GGC CGA GCT CTG CTA CAG TCC TGT ACA GTT CAG     663

Val Thr Gln Ala Pro Gly Gly Trp Glu Val Leu Ala Val
GTG ACA CAG GCC CCA GGA GGC TGG GAA GTC CTG GCT GTT     702

Val Val Pro Val Pro Pro Phe Thr Cys Leu Leu Arg Asp
GTG GTC CCT GTG CCC CCC TTT ACC TGC CTG CTC CGG GAC     741

Leu Val Pro Ala Thr Asn Tyr Ser Leu Arg Val Arg Cys
CTG GTG CCT GCC ACC AAC TAC AGC CTC AGG GTG CGC TGT     780
```

```
Ala Asn Ala Leu Gly Pro Ser Pro Tyr Ala Asp Trp Val
GCC AAT GCC TTG GGG CCC TCT CCC TAT GCT GAC TGG GTG    819

Pro Phe Gln Thr Lys Gly Leu Ala Pro Ala Ser Ala Pro
CCC TTT CAG ACC AAG GGT CTA GCC CCA GCC AGC GCT CCC    858

Gln Asn Leu His Ala Ile Arg Thr Asp Ser Gly Leu Ile
CAA AAC CTC CAT GCC ATC CGC ACA GAT TCA GGC CTC ATC    897

Leu Glu Trp Glu Glu Val Ile Pro Glu Ala Pro Leu Glu
TTG GAG TGG GAA GAA GTG ATC CCC GAG GCC CCT TTG GAA    936

Gly Pro Leu Gly Pro Tyr Lys Leu Ser Trp Val Gln Asp
GGC CCC CTG GGA CCC TAC AAA CTG TCC TGG GTT CAA GAC    975

Asn Gly Thr Gln Asp Glu Leu Thr Val Glu Gly Thr Arg
AAT GGA ACC CAG GAT GAG CTG ACA GTG GAG GGG ACC AGG   1014

Ala Asn Leu Thr Gly Trp Asp Pro Gln Lys Asp Leu Ile
GCC AAT TTG ACA GGC TGG GAT CCC CAA AAG GAC CTG ATC   1053

Val Arg Val Cys Val Ser Asn Ala Val Gly Cys Gly Pro
GTA CGT GTG TGC GTC TCC AAT GCA GTT GGC TGT GGA CCC   1092

Trp Ser Gln Pro Leu Val Val Ser Ser His Asp Arg Ala
TGG AGT CAG CCA CTG GTG GTC TCT TCT CAT GAC CGT GCA   1131

Gly Gln Gln Gly Pro Pro His Ser Arg
GGC CAG CAG GGC CCT CCT CAC AGC CGC                   1158
```

CLAIMS:

1. A mammalian receptor tyrosine kinase which is a developmental tyrosine kinase (Dtk) and which is expressed in multipotential haematopoietic cells, in embryonic stem cells, in brain tissue and in testis, but which is not expressed in mature lineage-restricted haematopoietic cells.

2. A receptor tyrosine kinase according to claim 1 that is murine Dtk having the amino acid sequence of SEQ ID NO 1, or a functional equivalent thereof.

3. A receptor tyrosine kinase according to claim 1 that is mature murine Dtk having the amino acid sequence of SEQ ID NO 2.

4. A receptor tyrosine kinase according to claim 1 that is human Dtk having the amino acid sequence of SEQ ID NO 3, or a functional equivalent thereof.

5. A receptor tyrosine kinase according to claim 1 that is mature human Dtk having the amino acid sequence of SEQ ID NO 4.

6. An extracellular receptor domain of a receptor tyrosine kinase according to claim 1.

7. An extracellular receptor domain which is the extracellular receptor domain of mature murine Dtk as defined in claim 3, or a functional equivalent thereof.

8. An extracellular receptor domain of a receptor tyrosine kinase having the amino acid sequence of SEQ ID NO 5.

9. An extracellular receptor domain which is the extracellular receptor domain of mature human Dtk as defined in claim 5, or a functional equivalent thereof.

10. An extracellular receptor domain of a receptor tyrosine kinase having the amino acid sequence of SEQ ID NO 6.

11. An extracellular receptor domain according to any one of claims 6 to 10 which is bound or attached to a support.

12. A soluble receptor comprising the extracellular receptor domain of a receptor tyrosine kinase according to any one of claims 1 to 5 lacking the transmembrane region and catalytic domain of said receptor tyrosine kinase.

13. A nucleic acid molecule encoding a receptor tyrosine kinase as defined in claim 1.

14. A nucleic acid molecule encoding murine Dtk or a functional equivalent thereof as defined in claim 2.

15. A nucleic acid molecule according to claim 14 which is DNA.

16. A DNA molecule according to claim 15 having the nucleotide sequence of SEQ ID NO 7.

17. A nucleic acid molecule encoding mature murine Dtk as defined in claim 3.

18. A nucleic acid molecule according to claim 17 which is DNA.

19. A DNA molecule according to claim 18 having the nucleotide sequence of SEQ ID NO 8.

20. A nucleic acid molecule encoding human Dtk or a functional equivalent thereof as defined in claim 4.

21. A nucleic acid molecule according to claim 20 which is DNA.

22. A DNA molecule according to claim 21 having the nucleotide sequence of SEQ ID NO 9.

23. A nucleic acid molecule encoding mature human Dtk as defined in claim 5.

24. A nucleic acid molecule according to claim 23 which is DNA.

25. A DNA molecule according to claim 24 having the nucleotide sequence of SEQ ID NO 10.

26. A nucleic acid molecule encoding an extracellular receptor domain as defined in claim 6.

27. A nucleic acid molecule encoding the extracellular receptor domain of murine Dtk or a functional equivalent thereof as defined in claim 7.

28. A nucleic acid molecule according to claim 27 which is DNA.

29. A DNA molecule according to claim 28 having the nucleotide sequence of SEQ ID NO 11.

30. A nucleic acid molecule encoding the extracellular receptor domain of human Dtk or a functional equivalent thereof as defined in claim 9.

31. A nucleic acid molecule according to claim 30 which is DNA.

32. A DNA molecule according to claim 31 having the nucleotide sequence of SEQ ID NO 12.

33. A vector including a DNA molecule as defined in claim 13.

34. A vector including a DNA molecule as defined in any one of claims 15, 16, 18 and 19.

35. A vector including a DNA molecule as defined in any one of claims 21, 22, 24 and 25.

36. A vector including a DNA molecule as defined in claim 28 or claim 29.

37. A vector including a DNA molecule as defined in claim 31 or claim 32.

38. A method of producing a receptor tyrosine kinase comprising the steps of:
  (a) culturing a host cell which has been transformed or transfected with a vector as claimed in any one of claims 33-35 to express the encoded receptor tyrosine kinase; and
  (b) recovering the expressed receptor tyrosine kinase.

39. A method of producing an extracellular receptor domain of a receptor tyrosine kinase comprising the steps of:
  (a) culturing a host cell which has been transformed or transfected with a vector as claimed in claim 36 or claim 37 to express the encoded extracellular receptor domain; and
  (b) recovering the expressed extracellular receptor domain.

40. A recombinant receptor tyrosine kinase which is the product of a method as defined in claim 38.

41. A recombinant extracellular receptor domain which is the product of a method as defined in claim 39.

42. A ligand that binds to a receptor tyrosine kinase as defined in claim 1.

43. A ligand that binds to a receptor tyrosine kinase as defined in claim 2.

44. A ligand that binds to a receptor tyrosine kinase as defined in claim 3.

45. A ligand that binds to a receptor tyrosine kinase as defined in claim 4.

46. A ligand that binds to a receptor tyrosine kinase as defined in claim 5.

47. A ligand that binds to an extracellular receptor domain of a receptor tyrosine kinase as defined in claim 6.

48. A ligand that binds to an extracellular receptor domain of a receptor tyrosine kinase as defined in claim 7.

49. A ligand that binds to an extracellular receptor domain of a receptor tyrosine kinase as defined in claim 8.

50. A ligand that binds to an extracellular receptor domain of a receptor tyrosine kinase as defined in claim 9.

51. A ligand that binds to an extracellular receptor domain of a receptor tyrosine kinase as claimed in claim 10.

52. A ligand that binds to an extracellular receptor domain of a receptor tyrosine kinase as claimed in claim 11.

53. A ligand that binds to a soluble receptor as defined in claim 12.

54. A ligand that binds to a receptor tyrosine kinase as claimed in claim 40.

55. A ligand that binds to an extracellular receptor domain as claimed in claim 41.

56. A ligand according to any one of claims 42-55 wherein the ligand stimulates the proliferation, differentiation and/or survival of cells which express a receptor tyrosine kinase according to claim 1.

57. A ligand according to any one of claims 42-55 wherein the ligand is antagonistic and at least partially blocks or inhibits the function of a receptor tyrosine kinase according to claim 1 through binding to said receptor.

58. A method of stimulating the proliferation, differentiation and/or survival of a cell expressing a receptor tyrosine kinase according to claim 1 comprising contacting the cell with a ligand according to claim 56.

59. A method according to claim 58 wherein the stimulation occurs *in vivo*.

60. A method according to claim 58 wherein the stimulation occurs *ex vivo*.

61. A method of inhibiting the function of a receptor tyrosine kinase according to claim 1 comprising contacting the receptor with a ligand according to claim 57.

62. A method according to claim 61 wherein the inhibition occurs *in vivo*.

63. A method according to claim 61 wherein the inhibition occurs *ex vivo*.

64. A method of treating a disease, syndrome or condition caused or mediated by an excess of a ligand as claimed in claim 56 comprising the step of contacting said excess of said ligand with an effective amount of a receptor tyrosine kinase according to any one of claims 1-5 and 40, an extracellular receptor domain according to any one of claims 6-11 and 41 or a soluble receptor according to claim 12.

65. A method of treating a disease, syndrome or condition caused or mediated by an excess of a ligand as defined in claim 57 comprising the step of contacting said excess of said ligand with an effective amount of a receptor tyrosine kinase according to any one of claims 1-5 and 40, an extracellular receptor domain according to any one of claims 6-11 and 41 or a soluble receptor according to claim 12.

66. A method of extracting a ligand as defined in claim 56 or claim 57 from a medium which may contain said ligand comprising the step of contacting said medium with a receptor tyrosine kinase according to any one of claims 1-5 and 40,

an extracellular receptor domain according to any one of claims 6-11 and 41 or a soluble receptor according to claim 12.

67. A method of isolating ligand(s) as defined in claim 56 or claim 57 from a medium which may contain said ligand(s), comprising the steps of:

(a) contacting said medium with an effective amount of a receptor tyrosine kinase according to any one of claims 1-5 and 40, an extracellular domain according to any one of claims 6-11 and 41 or a soluble receptor according to claim 12;

(b) detecting which ligand(s) bind to said tyrosine kinase receptor, extracellular receptor domain or soluble receptor; and

(c) isolating such bound ligand(s).

68. A ligand which is isolated by a method according to claim 67.

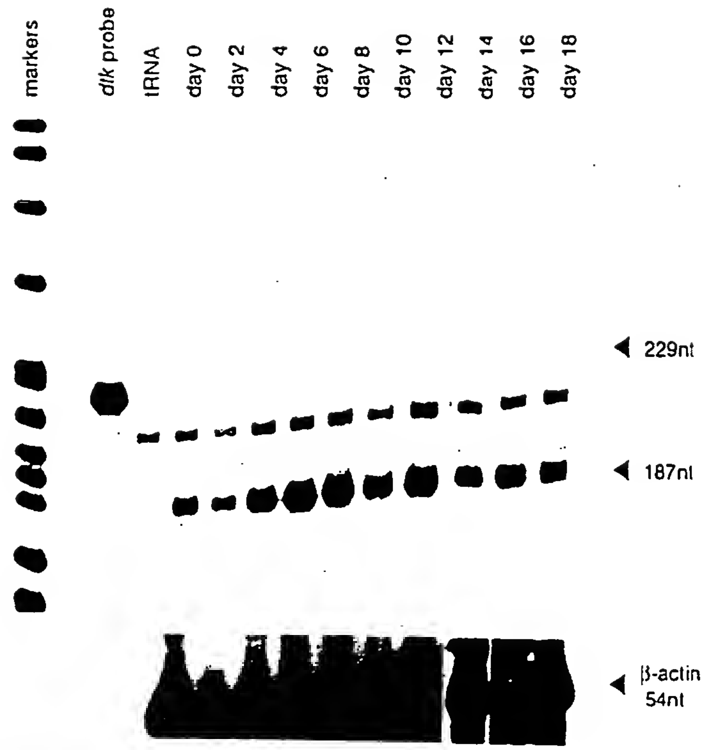

markers
dlk probe
tRNA
day 0
day 2
day 4
day 6
day 8
day 10
day 12
day 14
day 16
day 18
229nt
187nt
β-actin
54nt

FIG 1

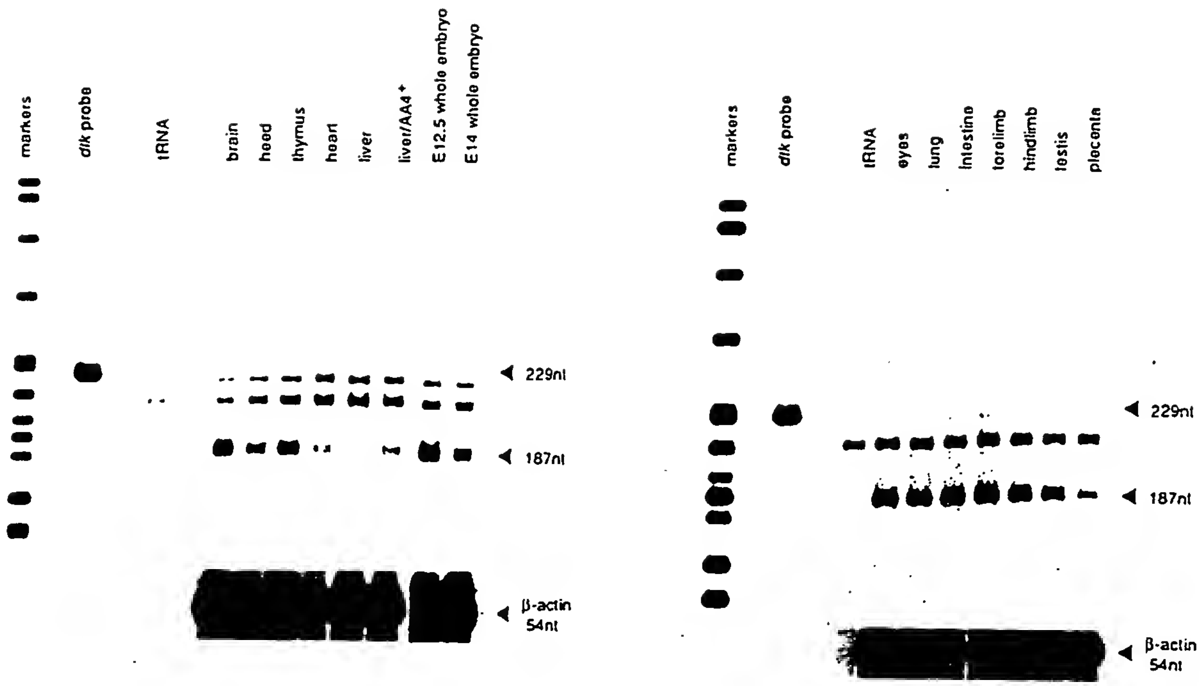

markers
dlk probe
tRNA
brain
head
thymus
heart
liver
liver/AA4+
E12.5 whole embryo
E14 whole embryo
229nt
187nt
β-actin
54nt
markers
dlk probe
tRNA
eyes
lung
intestine
forelimb
hindlimb
testis
placenta
229nt
187nt
β-actin
54nt

FIG 2

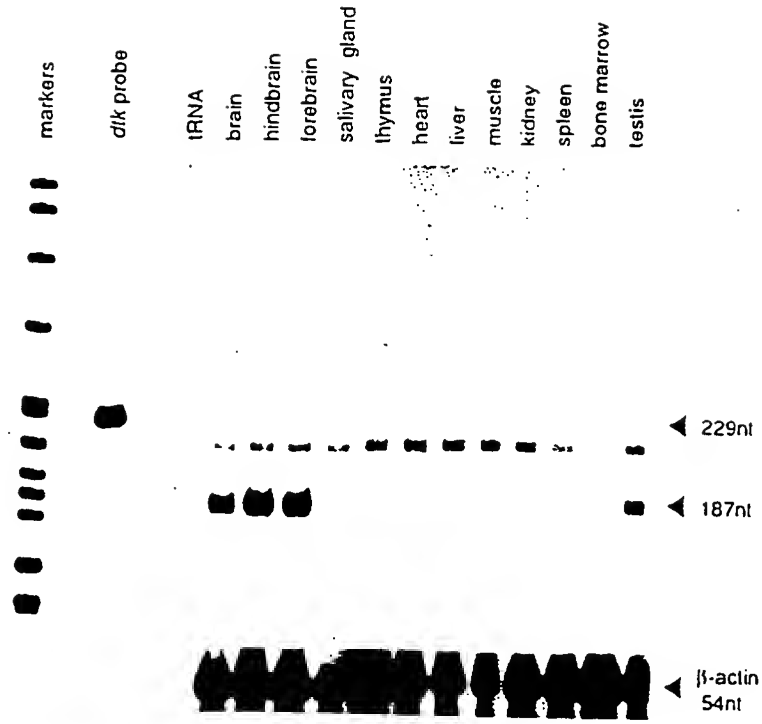

markers
dlk probe
tRNA
brain
hindbrain
forebrain
salivary gland
thymus
heart
liver
muscle
kidney
spleen
bone marrow
testis
229nt
187nt
β-actin
54nt

FIG 3

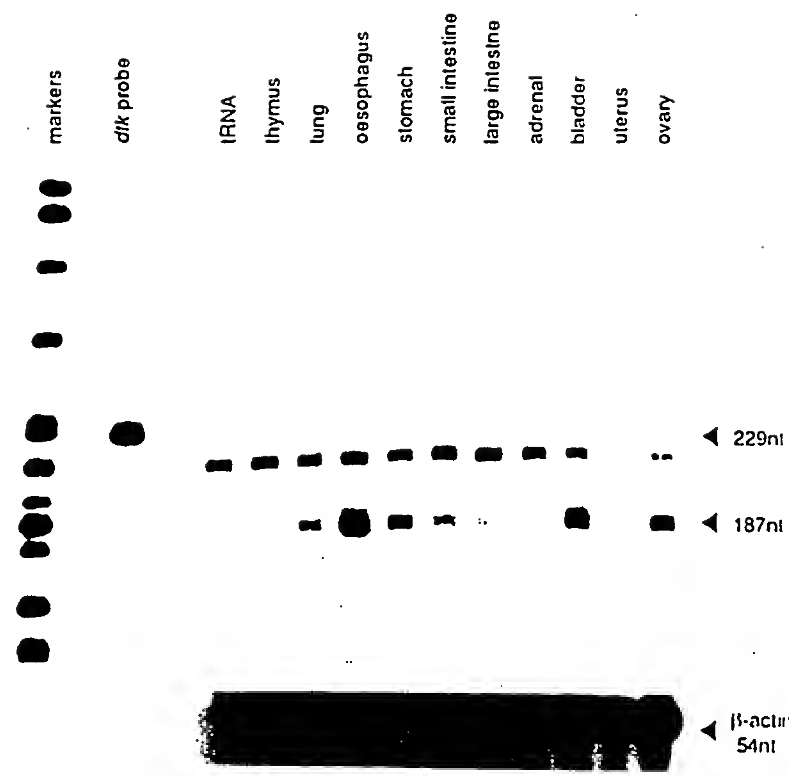

markers
dlk probe
tRNA
thymus
lung
oesophagus
stomach
small intestine
large intesine
adrenal
bladder
uterus
ovary
229nt
187nt
β-actin
54nt

FIG 4

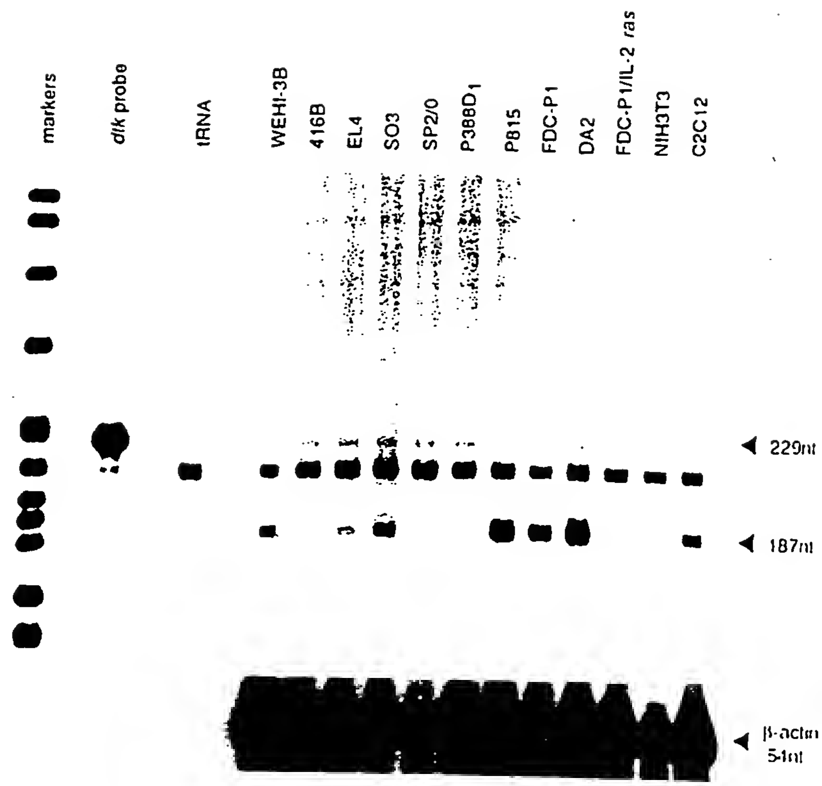

markers
dlk probe
tRNA
WEHI-3B
416B
EL4
SO3
SP2/0
P388D1
P815
FDC-P1
DA2
FDC-P1/IL-2 ras
NIH3T3
C2C12
229nt
187nt
β-actin
54nt

FIG 5

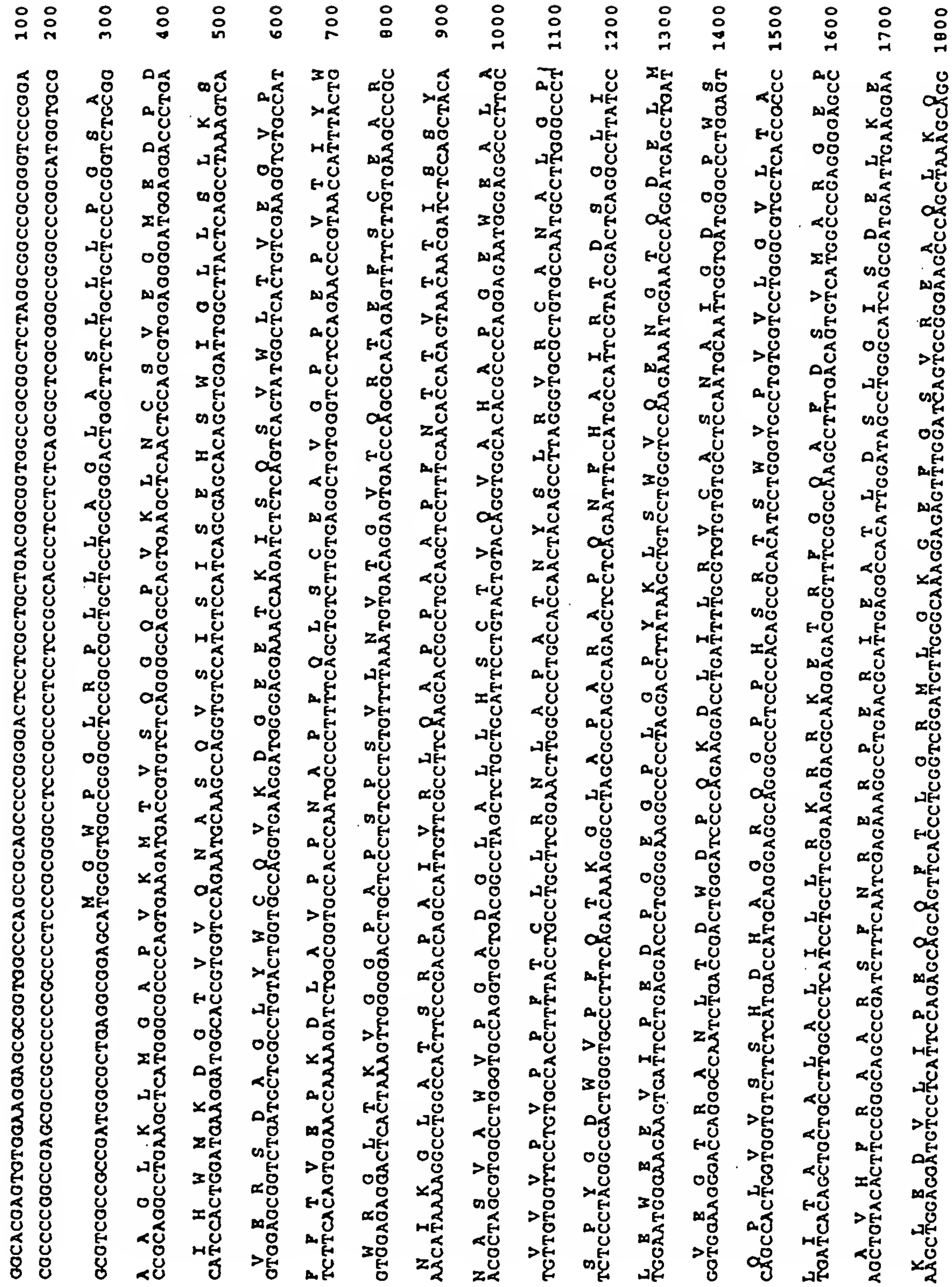


FIG 6

FIG 6 (cont)

```
E  D  G  S  F  V  K  V  A  V  K  M  L  K  A  D  I  I  A  S  S  D  I  E  E  F  L  R  E  A  A  C  M  K
AAGATGGCTCCTTCGTGAAAGTGGCAGTGAAGATGCTGAAAGCTGACATCATTGCCTCAAGCGACATAGAAGAGTTCCTCCGGGAAGCAGCTTGCATGAA 1900

E  F  D  H  P  H  V  A  K  L  V  G  V  S  L  R  S  R  A  K  G  R  L  P  I  P  M  V  I  L  P  F  M
GGAGTTTGACCATCCACACGTGGCCAAGCTTGTTGGGGTGAGCCTCCGGAGCAGGGCTAAAGGTCGTCTCCCCATTCCCATGGTCATCCTGCCCTTCATG 2000

K  H  G  D  L  H  A  F  L  L  A  S  R  I  G  E  N  P  F  N  L  P  L  Q  T  L  V  R  F  M  V  D  I
AAACATGGAGACTTGCACGCCTTTCTGCTCGCCTCCCGAATCGGGGAGAACCCTTTTAACCTGCCCCTGCAGACCCTGGTCCGGTTCATGGTGGACATTG 2100

A  C  G  M  E  Y  L  S  S  R  N  F  I  H  R  D  L  A  A  R  N  C  M  L  A  E  D  M  T  V  C  V  A  D
CCTGTGGCATGGAGTACCTGAGCTCCCGGAACTTCATCCACCGAGACCTAGCAGCTCGGAATTGCATGCTGGCCGAGGACATGACAGTGTGTGTGGCTGA 2200

F  G  L  S  R  K  I  Y  S  G  D  Y  Y  R  Q  G  C  A  S  K  L  P  V  K  W  L  A  L  E  S  L  A  D
TTTTGGACTCTCTCGGAAAATCTATAGCGGGGACTATTATCGTCAGGGCTGTGCCTCCAAATTGCCCGTCAAGTGGCTGGCCCTGGAGAGCTTGGCTGAC 2300

N  L  Y  T  V  H  S  D  V  W  A  F  G  V  T  M  W  E  I  M  T  R  G  Q  T  P  Y  A  G  I  E  N  A
AACTTGTATACTGTACACAGTGATGTGTGGGCCTTCGGGGTGACCATGTGGGAGATCATGACTCGTGGGCAGACGCCATATGCTGGCATTGAAAATGCTG 2400

E  I  Y  N  Y  L  I  G  G  N  R  L  K  Q  P  P  E  C  M  E  E  V  Y  D  L  M  Y  Q  C  W  S  A  D  P
AGATTTACAACTACCTCATCGGCGGGAACCGCCTGAAGCAGCCTCCGGAGTGCATGGAGGAAGTGTATGATCTCATGTACCAGTGCTGGAGCGCCGACCC 2500

K  Q  R  P  S  F  T  C  L  R  M  E  L  E  N  I  L  G  H  L  S  V  L  S  T  S  Q  D  P  L  Y  I  N
CAAGCAGCGCCCAAGCTTCACGTGTCTGCGAATGGAACTGGAGAACATTCTGGGCCACCTGTCTGTGCTGTCCACCAGCCAGGACCCCTTGTACATCAAC 2600

I  E  R  A  E  Q  P  T  E  S  G  S  P  E  L  H  C  G  E  R  S  S  S  E  A  G  D  G  S  G  V  G  A
ATTGAGAGAGCTGAGCAGCCTACTGAGAGTGGCAGCCCTGAGCTGCACTGTGGAGAGCGATCCAGCAGCGAGGCAGGGGACGGCAGTGGCGTGGGGGCAG 2700

V  G  G  I  P  S  D  S  R  Y  I  F  S  P  G  G  L  S  E  S  P  G  Q  L  E  Q  Q  P  E  S  P  L  N  E
TAGGTGGCATCCCCAGTGACTCTCGGTACATCTTCAGCCCCGGAGGGCTATCCGAGTCACCAGGGCAGCTGGAGCAGCAGCCAGAAAGCCCCCTCAATGA 2800

N  Q  R  L  L  L  L  Q  Q  G  L  L  P  H  S  S  C  *
GAACCAGAGGCTGTTGTTGCTGCAGCAAGGGCTACTGCCTCACAGTAGCTGTTAACCCTCAGGCAGAGGAAAGTTGGGGCCCCTGGCTCTGCTGACCACT 2900
GTGCTGCCTGACTAGGCCCAGTCTGATCACAGCCCAGGCAGCAAGGTATGGAGGCTCCTGTGGTAGCCCTCCCAAGCTGTGCTGGCGCCTGGACGGACCA 3000
AATTGCCCAATCCCAGTTCTTCCTGCAGCCGCTCTGGCCAGCCTGGCATCAGTTCAGGCCTTGGCTTAGAGGAGGTGAGCCAGAGCTGGTTGCCTGAATG 3100
CAGGCAGCTGGCAGGAGGGGAGGGTGGCTATGTTTCCATGGGTACCATGGGTGTGGATGGCAGTAAGGGAGGGTAGCAACAGCCCTGTGGGCCCCTACCC 3200
TCCTGGCTGAGCTGCTCCTACTTTAGTGCATGCTTGGAGCCGCCTGCAGCCTGGAACTCAGCACTGCCCACCACACTTGGGCCGAAATGCCAGGTTTGCC 3300
CCTCTTAAGTCACAAAGAGATGTCCATGTATTGTTCCCTTTTAGGTGATGATTAGGAAGGGATTGGCACACTTGGGTCCCTAAGCCCTATGGCAGGAAAT 3400
GGTGGGATATTCTCAGGTCTGAATCCTCATCATCTTCCTGATTCCCCACCCTGCAAAGGCCTGGAACTGGCTGTGGGGCTCTGAGGCATGCTGAAGGACA 3500
AAAGATTACAGAGATCCGACTTCAAAAGGCAGGGTCTGAGTCTGGCAGGTGGAGAGGTGCTAAGGGGCTGGCCCAGGAGTCAGGCATTTCAGGACCCCTC 3600
CAAGCTTCTACAGTCTGTCTGAGCATGCTACCAAGCCCCCAGATACCCCAAAACTAACAGAGGCAGTTTTGTCTGAGCCCAGCCCTCCCACATGATGACC 3700
CTTAGGTCTACCCTCCTCTCTAAATGGACATCCTCGTTTGTCCCAAGTCTCCAGAGAGACTACTGATGGCTGATGTGGGTAAGAAAAGTTCCAGGAACCA 3800
GGGCTGGGGTGGAACCAGGGCTGGGGTCGAGGCAGGCTCTTGGGCAGGCTCTTGCTGTTAGGAACATTTCTAAGCTATTAAGTTGCTGTTTCAAAACAAA 3900
TAAAATTGAAACATAAAG(A)n                                                                                3919
```

FIG 7

```
CATTAGATCTTTACATGAAAGTAAAATTTATAAGATTTCTAGAAAGTCAAAAGATGATAACTATTTCTTAGGATACTAAAAGCACTCACATTATAGAAAA  100
AAAATCAGTTAACTATACTCCACAAACATTAAAGGCTCCCTATAAAAAAACATTTTTAATAGGCAAGCCACAGAAAGGGCAAATATTAATAGTTTGCAAT  200
ACATATGTATGAAAAGGAATTGAATCTAGAATATTTAACAAAGCTTTACAACTCAAAAAATACAAAGAAAATATTTTTCTTCCAATTGGCAAATTACTTA  300
AACAGAACCTTCACAAAAGAAGATAAGAATGTTTAATAAACATTTGAAGCCATAATAATGACATCATTAGCCATGATGGAAATGCAAATTTAAGTACCAC  400
TTCACATCCACAAGAAAAAGATAAAAATAAAAGGACTGAGCTCACCAAACATTGGTGAGGATGTGGTAATACTGAAATTCTTGTACCGTGCTCCTGAGGG  500
TATAACATATTACAGGATTTTTTTGAAAACTAGTGGTTCCTTATAAACTTAATGCCCTGGCAACCTCACACCTATTTACTTAAGAATGAAAGGGCCCCGC  600
                                                                                          M  G  R  P  G  L  P  P  L  P
CCTCCTCCCTCCTCGCTCGCGGGCCGGGCCCGGCATGGTGCGGCGTCGCCGCCGATGGCGCTGAGGCGGAGCATGGGGCGGCCGGGGCTCCCGCCGCTGC  700
  L  P  P  P  P  R  L  G  L  L  L  A  E  S  A  A  A  G  L  K  L  M  G  A  P  V  K  L  T  V  S  Q  G
CGCTGCCGCCGCCACCGCGGCTCGGGCTGCTGCTGGCGGAGTCCGCCGCCGCAGGTCTGAAGCTCATGGGAGCCCCGGTGAAGCTGACAGTGTCTCAGGG  800
 Q  P  V  K  L  N  C  S  V  E  G  M  E  E  P  D  I  Q  W  V  K  D  G  A  V  V  Q  N  L  D  Q  L  Y
GCAGCCGGTGAAGCTCAACTGCAGTGTGGAGGGGATGGAGGAGCCTGACATCCAGTGGGTGAAGGATGGGGCTGTGGTCCAGAACTTGGACCAGTTGTAC  900
I  P  V  S  E  Q  H  W  I  G  F  L  S  L  K  S  V  E  R  S  D  A  G  R  Y  W  C  Q  V  E  D  G  G  E
ATCCCAGTCAGCGAGCAGCACTGGATCGGCTTCCTCAGCCTGAAGTCAGTGGAGCGCTCTGACGCCGGCCGGTACTGGTGCCAGGTGGAGGATGGGGGTG  1000
  T  E  I  S  Q  P  V  W  L  T  V  E  G  V  P  F  F  T  V  E  P  K  D  L  A  V  P  P  N  A  P  F  Q
AAACCGAGATCTCCCAGCCAGTGTGGCTCACGGTAGAAGGTGTGCCATTTTTCACAGTGGAGCCAAAAGATCTGGCAGTGCCACCCAATGCCCCTTTCCA  1100
 L  S  C  E  A  V  G  P  P  E  P  V  T  I  V  W  W  R  G  T  T  K  I  G  G  P  A  P  S  P  S  V  L
ACTGTCTTGTGAGGCTGTGGGTCCCCCTGAACCTGTTACCATTGTCTGGTGGAGAGGAACTACGAAGATCGGGGGACCCGCTCCCTCTCCATCTGTTTTA  1200
N  V  T  G  V  T  Q  S  T  M  F  S  C  E  A  H  N  L  K  G  L  A  S  S  R  T  A  T  V  H  L  Q  A  L
AATGTAACAGGGGTGACCCAGAGCACCATGTTTTCCTGTGAAGCTCACAACCTAAAAGGCCTGGCCTCTTCTCGCACAGCCACTGTTCACCTTCAAGCAC  1300
  P  A  A  P  F  N  I  T  V  T  K  L  S  S  S  N  A  S  V  A  W  M  P  G  A  D  G  R  A  L  L  Q  S
TGCCTGCAGCCCCCTTCAACATCACCGTGACAAAGCTTTCCAGCAGCAACGCTAGTGTGGCCTGGATGCCAGGTGCTGATGGCCGAGCTCTGCTACAGTC  1400
 C  T  V  Q  V  T  Q  A  P  G  G  W  E  V  L  A  V  V  V  P  V  P  P  F  T  C  L  L  R  D  L  V  P
CTGTACAGTTCAGGTGACACAGGCCCCAGGAGGCTGGGAAGTCCTGGCTGTTGTGGTCCCTGTGCCCCCCTTTACCTGCCTGCTCCGGGACCTGGTGCCT  1500
A  T  N  Y  S  L  R  V  R  C  A  N  A  L  G  P  S  P  Y  A  D  W  V  P  F  Q  T  K  G  L  A  P  A  S
GCCACCAACTACAGCCTCAGGGTGCGCTGTGCCAATGCCTTGGGGCCCTCTCCCTATGCTGACTGGGTGCCCTTTCAGACCAAGGGTCTAGCCCCAGCCA  1600
  A  P  Q  N  L  H  A  I  R  T  D  S  G  L  I  L  E  W  E  E  V  I  P  E  A  P  L  E  G  P  L  G  P
GCGCTCCCCAAAACCTCCATGCCATCCGCACAGATTCAGGCCTCATCTTGGAGTGGGAAGAAGTGATCCCCGAGGCCCCTTTGGAAGGCCCCCTGGGACC  1700
 Y  K  L  S  W  V  Q  D  N  G  T  Q  D  E  L  T  V  E  G  T  R  A  N  L  T  G  W  D  P  Q  K  D  L
CTACAAACTGTCCTGGGTTCAAGACAATGGAACCCAGGATGAGCTGACAGTGGAGGGGACCAGGGCCAATTTGACAGGCTGGGATCCCCAAAAGGACCTG  1800
I  V  R  V  C  V  S  N  A  V  G  C  G  P  W  S  Q  P  L  V  V  S  S  H  D  R  A  G  Q  Q  G  P  P  H
ATCGTACGTGTGTGCGTCTCCAATGCAGTTGGCTGTGGACCCTGGAGTCAGCCACTGGTGGTCTCTTCTCATGACCGTGCAGGCCAGCAGGGCCCTCCTC  1900
  S  R  T  S  W  V  P  V  V  L  G  V  L  T  A  L  V  T  A  A  A  L  A  L  I  L  L  R  K  R  R  K  E
ACAGCCGCACATCCTGGGTACCTGTGGTCCTTGGTGTGCTAACGGCCCTGGTGACGGCTGCTGCCCTGGCCCTCATCCTGCTTCGAAAGAGACGGAAAGA  2000
  T  R  F  G  Q  A  F  D  S  V  M  A  R  G  E  P  A  V  H  F  R  A  A  R  S  F  N  R  E  R  P  E  R
GACGCGGTTTGGGCAAGCCTTTGACAGTGTCATGGCCCGGGGAGAGCCAGCCGTTCACTTCCGGGCAGCCCGGTCCTTCAATCGAGAAAGGCCCGAGCGC  2100
```

FIG 7(cont)

```
I  E  A  T  L  D  S  L  G  I  S  D  E  L  K  E  K  L  E  D  V  L  I  P  E  Q  Q  F  T  L  G  R  M  L
ATCGAGGCCACATTGGACAGCTTGGGCATCAGCGATGAACTAAAGGAAAAACTGGAGGATGTGCTCATCCCAGAGCAGCAGTTCACCCTGGGCCGGATGT 2200

 G  K  G  E  F  G  S  V  R  E  A  Q  L  K  Q  E  D  G  S  F  V  K  V  A  V  K  M  L  K  A  D  I  I
TGGGCAAAGGAGAGTTTGGTTCAGTGCGGGAGGCCCAGCTGAAGCAAGAGGATGGCTCCTTTGTGAAAGTGGCTGTGAAGATGCTGAAAGCTGACATCAT 2300

 A  S  S  D  I  E  E  F  L  R  E  A  A  C  M  K  E  F  D  H  P  H  V  A  K  L  V  G  V  S  L  R  S
TGCCTCAAGCGACATTGAAGAGTTCCTCAGGGAAGCAGCTTGCATGAAGGAGTTTGACCATCCACACGTGGCCAAACTTGTTGGGGTAAGCCTCCGGAGC 2400

R  A  K  G  R  L  P  I  P  M  V  I  L  P  F  M  K  H  G  D  L  H  A  F  L  L  A  S  R  I  G  E  N  P
AGGGCTAAAGGCCGTCTCCCCATCCCCATGGTCATCTTGCCCTTCATGAAGCATGGGGACCTGCATGCCTTCCTGCTCGCCTCCCGGATTGGGGAGAACC 2500

 F  N  L  P  L  Q  T  L  I  R  F  M  V  D  I  A  C  G  M  E  Y  L  S  S  R  N  F  I  H  R  D  L  A
CCTTTAACCTACCCCTCCAGACCCTGATCCGGTTCATGGTGGACATTGCCTGCGGCATGGAGTACCTGAGCTCTCGGAACTTCATCCACCGAGACCTGGC 2600

 A  R  N  C  M  L  A  E  D  M  T  V  C  V  A  D  F  G  L  S  R  K  I  Y  S  G  D  Y  Y  R  Q  G  C
TGCTCGGAATTGCATGCTGGCAGAGGACATGACAGTGTGTGTGGCTGACTTCGGACTCTCCCGGAAGATCTACAGTGGGGACTACTATCGTCAAGGCTGT 2700

A  S  K  L  P  V  K  W  L  A  L  E  S  L  A  D  N  L  Y  T  V  Q  S  D  V  W  A  F  G  V  T  M  W  E
GCCTCCAAACTGCCTGTCAAGTGGCTGGCCCTGGAGAGCCTGGCCGACAACCTGTATACTGTGCAGAGTGACGTGTGGGCGTTCGGGGTGACCATGTGGG 2800

 I  M  T  R  G  Q  T  P  Y  A  G  I  E  N  A  E  I  Y  N  Y  L  I  G  G  N  R  L  K  Q  P  P  E  C
AGATCATGACACGTGGGCAGACGCCATATGCTGGCATCGAAAACGCTGAGATTTACAACTACCTCATTGGCGGGAACCGCCTGAAACAGCCTCCGGAGTG 2900

 M  E  D  V  Y  D  L  M  Y  Q  C  W  S  A  D  P  K  Q  R  P  S  F  T  C  L  R  M  E  L  E  N  I  L
TATGGAGGACGTGTATGATCTCATGTACCAGTGCTGGAGTGCTGACCCCAAGCAGCGCCCGAGCTTTACTTGTCTGCGAATGGAACTGGAGAACATCTTG 3000

G  Q  L  S  V  L  S  A  S  Q  D  P  L  Y  I  N  I  E  R  A  E  E  P  T  V  G  G  S  L  E  L  P  G  R
GGCCAGCTGTCTGTGCTATCTGCCAGCCAGGACCCCTTATACATCAACATCGAGAGAGCTGAGGAGCCCACTGTGGGAGGCAGCCTGGAGCTACCTGGCA 3100

 D  Q  P  Y  S  G  A  G  D  G  S  G  M  G  A  V  G  G  T  P  S  D  C  R  Y  I  L  T  P  G  G  L  A
GGGATCAGCCCTACAGTGGGGCTGGGGATGGCAGTGGCATGGGGGCAGTGGGTGGCACTCCCAGTGACTGTCGGTACATACTCACCCCCGGAGGGCTGGC 3200

 E  Q  P  G  Q  A  E  H  Q  P  E  S  P  L  N  E  T  Q  R  L  L  L  L  Q  Q  G  L  L  P  H  S  S  C
TGAGCAGCCAGGGCAGGCAGAGCACCAGCCAGAGAGTCCCCTCAATGAGACACAGAGGCTTTTGCTGCTGCAGCAAGGGCTACTGCCACACAGTAGCTGT 3300
•
TAGCCCACAGGCAGAGGGCATCGGGGCCATTTGGCCGGCTCTGGTGGCCACTGAGCTGGCTGACTAAGCCCCGTCTGACCCCAGCCCAGACAGCAAGGTG 3400
TGGAGGCTCCTGTGGTAGTCCTCCCAAGCTGTGCTGGGAAGCCCGGACTGACCAAATCACCCAATCCCAGTTCTTCCTGCAACCACTCTGTGGCCAGCCT 3500
GGCATCAGTTTAGGCCTTGGCTTGATGGAAGTGGGCCAGTCCTGGTTGTCTGAACCCAGGCAGCTGGCAGGAGTGGGGTGGTTATGTTTCCATGGTTACC 3600
ATGGGTGTGGATGGCAGTGTGGGGAGGGCAGGTCCAGCTCTGTGGGCCCTACCCTCCTGCTGAGCTGCCCCTGCTGCTTAAGTGCATGCATTGAGCTGCC 3700
TCCAGCCTGGTGGCCCAGCTATTACCACACTTGGGGTTTAAATATCCAGGTGTGCCCCTCCAAGTCAGAAAGAGATGTCCTTGTAATATTCCCTTTTAGG 3800
TGAGGGTTGGTAAGGGGTTGGTATCTCAGGTCTGAATCTTCACCATCTTTCTGATTCCGCACCCTGCCTACGCCAGGAGAAGTTGAGGGGAGCATGCTTC 3900
CCTGCAGCTGACCGGGTCACACAAAGGCATGCTGGAGTACCCAGCCTATCAGGTGCCCCTCTTCCAAAGGCAGCGTGCCGAGCCAGCAAGAGGAAGGGGT 4000
GCTGTGAGGCTTGCCCAGGAGCAAGTGAGGCCGGAGAGGAGTTCAGGAACCCTTCTCCATACCCACAATCTGAGCACGCTACCAAATCTCAAAATATCCT 4100
AAGACTAACAAAGGCAGCTGTGTCTGAGCCCAACCCTTCTAAACGGTGACCTTTAGTGCCAACTTCCCCTCTAACTGGACAGCCTCTTCTGTCCCAAGTC 4200
TCCAGAGAGAAATCAGGCCTGATGAGGGGGAATTCCTGGAACCTGGACCCCAGCCTTGGTGGGGGAGCCTCTGGAATGCATGGGGCGGGTCCTAGCTGTT 4300
AGGGACATTTCCAAGCTGTTAGTTGCTGTTTAAAATAGAAATAAAATTGAAGACTAAAGACCT(A)n                                   4364
```